11-1-1995

An Investigation into world wide web publishing with the hypertext markup language

Eric Joseph Cohen

An Investigation into World Wide Web Publishing with the Hypertext Markup Language

by Eric Joseph Cohen

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in the
School of Printing Management and Sciences in the
College of Imaging Arts and Sciences of the
Rochester Institute of Technology

November 1995

Thesis Advisor: Professor Frank Romano

School of Printing Management and Sciences
Rochester Institute of Technology
Rochester, New York

Certificate of Approval

Master's Thesis

This is to certify that the Master's Thesis of

Eric Joseph Cohen

With a major in *Graphic Arts Publishing*
has been approved by the Thesis Committee as satisfactory
for the thesis requirement for the Master of Science degree
at the convocation of

November 1995

Thesis Committee:

Frank Romano
Thesis Advisor

Marie Freckleton
Graduate Program Coordinator

C. Harold Goffin
Director or Designate

Title of Thesis:

An Investigation into World Wide Web Publishing with the Hypertext Markup Language

September 12, 1995

Acknowledgements

Many thanks to everyone who contributed time and effort to this project. Thanks are especially due to Professor Frank Romano for his inspiration, patience, and careful, steadfast guidance throughout the course of this thesis project.

Most importantly, I would like to express my gratitude towards my family. This project would never have been realized without their encouragement and support. Thank you with all my love.

Table of Contents

List of Tables vi

List of Figures vii

Abstract . . . . . viii–ix

Chapter 1
Introduction 1
Reasons for Interest 5
End Notes 7

Chapter 2
Theoretical Bases of the Study 8
Data Size 8
Physical Size 9
Data Formats 9
Text-Only Users 10
End Notes 11

Chapter 3
Review of Literature in the Field of Study 12
End Notes 15

Chapter 4
Statement of Project Goals 16

Chapter 5
Methodology 17
Software 18
Hardware 18

Chapter 6
Results 19
Compiled Information 19
HTML Programming 19
Style Issues 23
Implementing Mixed Media Files 25
Inline versus external images 25
Inline images 25

External images 27
Sound 28
Video 29
Server Issues 31
HTTP 32

Chapter 7
Summary and Conclusions 34
Data Size 34
Physical Size 35
Size and Format Indicators 36
Text-Only Users 37
Recommended areas for further study 40

Bibliography 41–44

Appendix A
HTML Tag Reference 46–51

Appendix B
Web Site's HTML Documents 52–62

Appendix C
Electronic Publisher's Guide to the WWW 63–100

List of Tables

Table 1
External Image formats and extensions 28

Table 2
Sound formats and extensions 29

Table 3
Video formats and extensions 30

List of Figures

Figure 1
URL Description 3

Figure 2
Home Page Source Code 21

Figure 3
Graphical Representation of Home Page 22

Figure 4
Mosaic Representation 24

Figure 5
Netscape Representation 24

Figure 6
Text-only Representation 38

Abstract

The purpose of this thesis project was to test and to demonstrate the World Wide Web as a publishing vehicle by creating a Web presence for the School of Printing Management and Sciences. In order to reach this goal, a full understanding of the Hypertext Markup Language must first be realized. Once this is accomplished, issues regarding integration of mixed-media elements within an HTML document were investigated. Once a prototype of the HTML document was accomplished, the mixed-media elements were tested and evaluated for proper integration and contextual cohesiveness. Many issues regarding implementation of mixed-media elements, such as file size and file format were addressed upon testing.

One of the additional goals of this project is a comprehensive description of the methodology for creating and maintaining a World Wide Web publishing presence. This addresses: navigational software, structuring HTML documents, hypertext linking, HTML style issues and limitations, effective integration of mixed-media elements, inline and external image issues, testing documents, advertising documents, strategies for determining proper file sizes and formats of mixed-media elements, integrating supplemental programs, World Wide Web Server issues, installing HTML and mixed-media files onto a World Wide Web Server, etc.

The Web site located at `(http://www.rit.edu/~spms)` served as the vehicle for the investigation. Results of the study revealed the issues of providing data that services users across a wide range of computer systems, with different bandwidth restrictions, utilizing a myriad of computer software. Specific standards apply to

alleviate much of the guesswork, however, publishing on the Internet remains to be as challenging as it is rewarding.

The Web's format and the opportunity to reach millions of potential customers is creating new types of publishing ventures in true "gold-rush" fashion. The Web is being touted as the fourth medium, and some suggest it will have as great an impact on society as print, radio and television. The growth of the Web is explosive and will assuredly continue to blossom.

Upon completion of this study, the author remains skeptical whether the World Wide Web is the medium of the future. It has, however, created a trend which will forever reshape the publishing world and the way information seekers receive their data. Publishing will change from a commodity based market where prices are based upon cost, and shift to a service market where prices are based upon the value of the information. Each reader requiring selected information tailored to their specific choice will pay for what they select—no more paying for an entire magazine or newspaper and reading only one article. The future of information dissemination is electronic, interactive and selective. Whether the delivery mechanism will be the World Wide Web remains to be seen.

Chapter 1: Introduction

The development and explosive growth of the World Wide Web has necessitated a need to develop new tools and techniques designed to harness the technology involved in Internet publishing. Currently, a World Wide Web publisher is faced with a myriad of hardware and software considerations for the development and maintenance of a specified World Wide Web site. This thesis project will provide a potential World Wide Web publisher with the proper knowledge and guidelines for publishing on the World Wide Web. This was accomplished by creation of a World Wide Web presence for the School of Printing Management and Sciences and then authoring a manual detailing the step-by-step approach a potential electronic publisher must follow to successfully publish information on the World Wide Web.

The World Wide Web is the world's largest "document"—with thousands and thousands of text, graphics, audio and video files interlinked throughout the world on connected servers. Developed in 1991 at CERN, the European Particle Physics Laboratory in Geneva, Switzerland, it was originally conceived as a hypertext exchange system to allow scientists to share their findings with other scientists over the Internet. The World Wide Web was met with great enthusiasm and quickly grew beyond the bounds of CERN. Computer enthusiasts of all kinds started developing World Wide Web applications and promoting the World Wide Web across the Internet. It did not take long before all types of hypermedia became available on the World Wide Web for Internet enthusiasts to explore. The World Wide Web's use and popularity began to blossom, suggesting the future of hypermedia would grow.

Unfortunately, at that point, the World Wide Web was still difficult for anyone but Internet professionals to navigate. This changed in June of 1993 when the National Center for Supercomputing Applications (NCSA) introduced a simple and powerful graphical user interface for World Wide Web navigating. The software is called Mosaic, which can be retrieved from the Internet for free, and is currently revolutionizing the World Wide Web and bringing it to the masses. In June of 1993 a software robot was sent through the World Wide Web to count how many sites existed. The robot found around one hundred such sites. In May of 1994, a similar robot logged onto 3,800 sites.[1] This significant increase in new sites is due almost entirely to Mosaic.

The structure of the World Wide Web allows for this great influx of new sites and the millions of World Wide Web users visiting these sites. The World Wide Web is based on the Hypertext Transport Protocol (HTTP), a protocol designed for sharing documents via a client server model. Hypertext Markup Language (HTML) is simply the language that all World Wide Web documents speak. It is a set of tags that tell the navigational software how to display text, hypertext links, images, and movie clips. HTML was purposely kept simple; there are only a few tags for text markup, allowing navigational software to be minimal and thus inexpensive to write. Anyone can download Mosaic for free; and it was actually written by students.

Not only does Mosaic bring World Wide Web exploration to the masses, but the simple tagging schemes of HTML allow novice World Wide Web users to create their own documents. These homemade documents or home pages, can now be

hyperlinked to any other home pages residing on a World Wide Web server via their Uniform Resource Locator (URL). A URL is a system that the World Wide Web uses to standardize the addresses of various Internet services. A URL "expresses the address of a resource and the method by which it can be accessed."[2] In short, a URL provides location information. A typical URL looks like this:

```
http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/MosaicHome.html
```

PROTOCOL	`http:`
SERVER	`//www.ncsa.uiuc.edu`
PATHNAME	`/SDG/Software/Mosaic/MosaicHome.html`

Figure 1
URL Description

Within the URL, text are components that specify the protocol, server, and pathname of an item. The protocol is followed by a colon (http:), the server is preceded by two slashes (`//www.ncsa.uiuc.edu`), and each segment of the pathname is preceded by a single slash (`/SDG/Software/Mosaic/MosaicHome.html`).

The first component, the protocol, identifies a manner for interpreting computer information. Many Internet pages use HTTP (short for HyperText Transfer Protocol). Other common protocols include file (also known as ftp, short for File Transfer Protocol), news (the protocol used by Usenet news groups), and gopher (an alternative transfer protocol).[3]

The second component, the server, identifies the computer system that stores the information you seek. Each server on the Internet has a unique address name whose text refers to the organization maintaining the server.[4]

The last component, the pathname, identifies the location of an item on the server. For example, a pathname might identify a page by specifying the name of the file comprising the page (`/MosaicHome.html`) as well as the name of one or more folders that store a file (`/Mosaic`).[5]

Since Mosaic is such a "bare-bones" software package, supplemental programs are needed to achieve full benefit of the World Wide Web's hypermedia capabilities. These programs reside on one's hard disk and are referred to as Helper programs. This scheme allows Mosaic to keep its focus on its business, while letting other sophisticated viewing and listening applications handle the media presentations. Depending on the platform running, it obviously dictates which supplemental programs are used. Typical programs running on a Macintosh are SimplePlayer for viewing QuickTime clips, Sparkle for viewing MPEG clips, SoundMachine for AU sounds, and JPEGView for viewing GIF/JPEG images.

HTML offers great flexibility in its tagging schemes. Tags are guidelines for how a software navigator should display the material. Some tags are required by all HTML software navigators, while other tags are optional. All World Wide Web software navigators are different. Some, like Mosaic, are graphical, while others are text based. HTML is only a set of guidelines for displaying the documents—not everyone has to have the same capabilities to access World Wide Web informa-

tion, such as text-only users. Once the HTML document arrives at its destination, the software navigator can use whatever tags are appropriate to it and ignore the rest. For instance, Lynx, another popular World Wide Web browser, supports text only. It therefore ignores the HTML style information that's not relevant to it. Mosaic, on the other hand, presents documents in all their multimedia glory. This flexibility allows users to change the way World Wide Web documents look on their computer system.

The overwhelming success of NCSA Mosaic has brought an onslaught of new software navigators, all vying for their share of the marketplace. Currently, Netscape Communication's World Wide Web browser, called Netscape Navigator, is proving to overtake NCSA Mosaic as the defacto World Wide Web navigational software. Netscape Navigator supports later releases of HTML, therefore giving the designer more control of the style of their document.

Hypermedia is a major attribute of the Internet, but all file retrieval capabilities are at the mercy of bandwidth. Direct or dedicated connections to the Internet are suggested to take as much advantage of the World Wide Web's media delivery capabilities. Using SLIP or PPP dial-up accounts to retrieve QuickTime or MPEG video clips may be frustrating. SLIP/PPP connections use high-speed modems (14,400 baud) and special software to connect a home computer to the Internet.

Reasons for Interest

"The pen is no longer the only writing implement mightier than the sword, and freedom of the press no longer requires a press. The mighty pen and grinding

printing press are being challenged by something new: a quirky language called HTML (Hypertext Markup Language) and a paperless printer known as the World Wide Web."[6]

The printed medium is a wonderful source for disseminating information and will assuredly exist forever. However, more glamorous, cheaper, quicker, and environmentally sound methods of communicating the same information are hitting the mainstream. One does not have to be a printer. One should be an Electronic Publisher, an information gatherer, transformer, and deliverer. This outlook creates a wealth of new opportunities and challenges. Repackaging information in new ways creates opportunities, and develops new markets, and new audiences. This thesis project will encompass the aforementioned opportunities and challenges.

Print, interactive multimedia, and online publishing are solid markets of opportunity, but still relatively saturated. However, collectively, these media preview the future of information delivery. The opportunities are endless, limited only by the imagination of human endeavor, and bandwidth. Currently, the World Wide Web allows all these individual information technologies to merge, thus creating a new paradigm for information delivery.

End Notes for Chapter One

[1]Branwyn, G. *Mosaic Quick Tour.* (Chapel Hill: Ventana Press, 1994) 9.

[2]Ford, A. *Spinning the Web.* (London: Thomson Publishing, 1995) 34.

[3]Ibid.

[4]Ibid.

[5]Ibid., 35.

[6]Savetz, M. "The Medium is the Matrix," *Internet World,* April 1995, page 70.

Chapter 2: Theoretical Basis of the Study

The effective integration of multimedia elements such as graphics, audio, and video into a World Wide Web site can be just as important as informative and well-organized text. When done properly, the incorporation of multimedia into a site can substantially increase its utility as well as it's esthetic appeal. However, poorly implemented multimedia can just as easily diminish a site's usefulness. When dealing with multimedia elements, a number of issues must be carefully considered. "In stark contrast to traditional ink on paper, publishing on the World Wide Web can be described as instantaneous communication, distributed worldwide to millions of readers in content-rich, colorful, highly interactive format, produced for a unit cost too cheap to meter."[1]

Data Size

"The learning curve associated with setting up and maintaining a WWW site is progressive. As a WWW site administrator or Webmaster, becomes comfortable with the site's basic functionality, more advanced features can be added. However, implementing these features generally requires greater levels of technical expertise."[2] Implementing mixed-media elements into a World Wide Web document increases the overall file size of that document.[3] The more elements that are included, the higher the file size. This is the single largest issue confronting online information providers to date. The technology allows for inclusion of video, sound, and large graphics, but most connections are ill equipped to download the data within a reasonable time frame.

Certain navigational software tackles the large graphics issue head-on. Users can specify parameters whether to allow the graphics to be viewed or not. If a user chooses not to show the graphics, a small icon appears notifying that a graphic exists. This significantly reduces the retrieval time.

Audio and video files are relatively large[4], thus emphasizing the file size issue. Content providers who implement audio and video files within a document must seriously consider the end viewer when deciding the size of such elements. It is a general rule that potential downloaders are warned of any files over 100k.[5]

Physical Size

Another important consideration for graphics files is their physical size (height and width). Although it is directly related to data size, it is important to keep the physical size within reasonable limits so it looks acceptable on most viewers. Similarly, it is important to use reasonable length video and audio clips[6] (less than one minute). Various software packages offer techniques on how to address this problem.

Data Formats

World Wide Web publishers create their multimedia elements across the gamut of computer platforms utilizing hundreds of software packages. The resulting effect is a myriad of possible file formats facing the World Wide Web publisher. Also, not all navigational software and supplemental programs are suited to handle the different file formats. Certain standards exist that attempt to tackle this issue.

Text-Only Users

Many World Wide Web users still use text-only navigational software. In fact, until two years ago, text-only navigational software was all that existed.[7] It is frustrating for a text-only World Wide Web user to try to access information, only to find out that the particular site is only formatted for graphical navigational software. These users are often ignored, reducing the value of a World Wide Web site. Strategies exist on how to best orient a World Wide Web site for all types of navigational software.

The World Wide Web ". . . has great strengths in support for graphics and hypertext links among documents, it has needed better security to enable financial transactions to take place safely. It has cried for a richer markup language to handle tables and style sheets. And it has required greater efficiency in expediting transactions."[8] The Web is touted as the next great information delivery vehicle, there are still many shortcomings that must be solved before it can become a mainstream approach.

End Notes for Chapter Two

[1]Haiman, S. The World Wide Wonder. *Color Publishing,* March/April 1995. 34.

[2]Cutler, M. "Setting Up Your Own Web Site." *Web Watch,* April 1995. 1.

[3]Lemay, L. *Teach Yourself Web Publishing with HTML in a Week.* (Indianapolis: Sams Publishing, 1995) 183.

[4]Ibid., 193

[5]Ibid., 194

[6]Ibid.

[7]Wiggins, R. "Webolution," *Internet World.* April 1995. 33.

[8]Editor. (1994, October 10). "Web Publishing: Pieces Falling into Place." *Seybold Report on Desktop Publishing,* 10 October 1994. 3.

Chapter 3: Review of Literature in the Field of Study

Haiman, Steven. "The World Wide Wonder." ***Color Publishing.*** **March/April: 34-39**

This article details the World Wide Web from a publisher's perspective. It describes the inherent pitfalls and roadblocks traditional print publishers encounter before their information reaches the public. "In stark contrast to traditional ink on paper, publishing on the World Wide Web can be described as instantaneous communication, distributed worldwide to millions of readers in content-rich, colorful, highly interactive format, produced for a unit cost too cheap to meter."[1] It should be noted that an article such as this appearing in a journal targeted for commercial printers, emphasizes the inroads online publishing is enjoying.

Weiss, Aaron. "Hop, Skip, and Jump." ***Internet World.*** **April 95: 41-44**

The proliferation of the World Wide Web and its ability to let individuals contribute information on it, has resulted in an explosion of new sites offering a wealth of diversified information. With so many new sites appearing on the World Wide Web, one needs a strategy of navigating and selecting the sites one so desires. This article suggests certain World Wide Web sites which act as a jumpstation for Internet travel. It is extremely easy to get lost within the World Wide Web, and these sites provide a safe haven for the lost to return. Certain sites also offer recommendations or criticisms for other sites worth investigating or avoiding.

Lemay, Laura. "Teach Yourself Web Publishing with HTML in a Week." Indianapolis: Sams Publishing, 1994.

This book provides a resource for the basic principles and theory involved in publishing on the Web. It covers issues such as HTML, HTTP, hypertext linking, Helper programs, etc. The downside of this book is it covers the issues in a generalized fashion instead of offering specifics. Therefore, it is not a reasonable resource or "how-to" manual. Also, since its release, greater, easier, and more efficient means of publishing on the Web have surfaced, thus rendering this book mostly obsolete.

Cutler, M. "Setting Up Your Own Web Site." *Web Watch.* April 95: 1.

"The learning curve associated with setting up and maintaining a WWW site is progressive. As a WWW site administrator or Webmaster, becomes comfortable with the site's basic functionality, more advanced features can be added. However, implementing these features generally requires greater levels of technical expertise."[2] Issues covered in this article include: network access, host computer hardware, server software, information content and structure, and ongoing maintenance. Overall, this article serves as a decent resource by forewarning a potential Web publisher on some of the obstacles that need to be addressed for successful implementation of a Web site.

Editor. "Web Publishing: Pieces Falling into Place." *Seybold Report on Desktop Publishing.* 10 October 1994: 3-5.

Although the World Wide Web ". . . has great strengths in support for graphics and hypertext links among documents, it has needed better security to enable financial transactions to take place safely. It has cried for a richer markup lan-

guage to handle tables and style sheets. And it has required greater efficiency in expediting transactions."[3] Although the Web is touted as the next great information delivery vehicle, there are still many shortcomings that must be solved before it can hit mainstream. This article addresses these shortcomings and what individuals, consortiums, and companies are doing to overcome and solve these shortcomings. Most companies that are involved in the (r)evolution of the Web, plan to use the Internet as a means of commerce.

End Notes for Chapter Three

[1]Haiman, S. The World Wide Wonder. *Color Publishing,* March/April 1995. 34.

[2]Cutler, M. "Setting Up Your Own Web Site." *Web Watch,* April 1995. 1.

[3]Editor. (1994, October 10). "Web Publishing: Pieces Falling into Place." *Seybold Report on Desktop Publishing,* 10 October 1994. 3.

Chapter 4: Statement of Project Goals

The goal of this thesis project is to test and demonstrate the World Wide Web as a publishing vehicle by creating a World Wide Web presence for the School of Printing Management and Sciences. In order to reach this goal, a full understanding of the Hypertext Markup Language must first be realized. Once this is accomplished, issues regarding integration of mixed-media elements within an HTML document can be investigated. Once a prototype of the HTML document is accomplished, the mixed-media elements were tested and evaluated for proper integration and contextual cohesiveness. Many issues regarding implementation of mixed-media elements, such as file size and format were addressed upon testing.

One of the additional goals of this project was a comprehensive description of the methodology for creating and maintaining a World Wide Web publishing presence. This addresses: navigational software, structuring HTML documents, hypertext linking, HTML style issues and limitations, effective integration of mixed-media elements, inline and external image issues, testing documents, advertising documents, strategies for determining proper file sizes and formats of mixed-media elements, integrating supplemental programs, World Wide Web Server issues, installing HTML and mixed-media files onto a Web Server, etc.

Chapter Five—Methodology

- Compiled information suitable for inclusion within a World Wide Web site for the School of Printing Management and Sciences. Information includes: program descriptions (Graduate & Undergraduate), course descriptions (Graduate & Undergraduate), the Print RIT Journal in a Portable Document Format, student papers, faculty contributions, related hypertext links (printing & publishing companies).

- Programmed the Hypertext Markup Language with the aid of online tutorials and developed a working prototype.

- Developed proper styles governing HTML documents.

- Implemented mixed-media elements within an HTML document.

- Compiled documents and transferred them into a UNIX server for world accessibility.

- Evaluated HTML documents for quality, transfer rates, size, formats and determine which combination works best.

- Authored a manual detailing the process of publishing on the World Wide Web utilizing all that has been learned. Issues covered within the manual include: navigational software, structuring HTML documents, hypertext

linking, HTML style issues and limitations, effective integration of mixed-media elements, inline and external image issues, testing documents, advertising documents, strategies for determining proper file sizes and formats of mixed-media elements, integrating supplemental programs, World Wide Web Server issues, installing HTML and mixed-media files onto a World Wide Web Server, etc.

The following equipment , software, and facilities were implemented to facilitate the completion of this thesis project.

Software

- Adobe Acrobat 2.0
- Adobe Photoshop 2.51
- A simple text editor (e.g. SimpleText)
- Pacerlink 5.3
- SoundEdit Pro 1.0.5
- FusionRecorder 1.0.2
- JPEGView 3.3
- MoviePlayer
- QuarkXPress 3.31
- Transparency 1.0
- Gif Converter 2.3
- Sparkle 1.71
- SoundMachine 2.1
- Netscape 1.1
- NCSA Mosaic 3.0
- Fetch 2.11

Hardware

- Macintosh computer capable of running the above software
- Internet connection
- Desktop flatbed scanner
- Video camera/VCR

Chapter 6: Results

Compiled information

The nature of this thesis project was to test and evaluate specific elements of publishing on the World Wide Web. the first step was to create a Web presence for the School of Printing Management and Sciences that was to act as the vehicle for testing and evaluation. The present address is `(http://www.rit.edu/~spms)`. The creation of a Web presence for an organization requires a designate to compile information suitable for inclusion within the site.

Material deemed pertinent for this experiment included: program descriptions (Graduate & Undergraduate), course descriptions (Graduate & Undergraduate), the Print RIT Journal in a Portable Document Format (PDF), a student paper, a faculty contribution, and related hypertext links (printing & publishing companies). A section was included entitled "Grab Bag", intentionally set up for the inclusion of future material.

The data was supplied in electronic form in almost all the cases resulting in a streamlined data exchange. Minimal text input was required for the introduction and hypertext descriptions. The PDF documents were left completely untouched as to maintain the integrity of the document that the author intended.

HTML programming

Once the relevant information was compiled, it had to be converted into HTML documents so the Web browser could interpret them. The Hypertext Markup Language is designed to specify the logical organization of a document, with

important hypertext extensions. It is not a WYSIWYG word processor such as Microsoft Word or WordPerfect. This is because the same document may be viewed by many different browsers, of very different abilities. Thus, for example, HTML allows you to mark titles or paragraph marks, and then leaves the interpretation of these marked elements up to the browser. For example one browser may indent the beginning of a paragraph, while another may only leave a blank line.

HTML instructions are called elements. These can be divided into two broad categories—those that define how the body of the document is to be displayed by the browser, and those that define information about the document, such as the title or relationships to other documents. Elements are denoted by the tag `<element_name>`. This is simply the element name surrounded by left and right angle brackets. Most elements mark blocks of the document for a particular purpose or for formatting: the above `<element_name>` tag marks the beginning of such as section. The end of this section would then be marked by the ending tag `</element_name>`. A reference to the most used HTML tags and their functions is listed in **Appendix A**.

HTML documents are structured into two parts, the head, and the body. The head contains information about the document that is not generally displayed with the document, such as its title. The body contains the body of the text, and is where the document material is to be displayed. **Figure 2** displays the HTML source code for the Web site's Home Page. The tags are in bold to emphasize their placement. **Figure 3** shows the graphical representation of the page as displayed by the Netscape browser. A complete reference for the documents within the Web site, their source code and graphical representations reside in **Appendix B**.

```
<HTML><HEAD><TITLE>School of Printing Management and Sciences</TITLE></HEAD><BODY>
<H3><A HREF="http://www.rit.edu/">ROCHESTER INSTITUTE OF TECHNOLOGY</A></H3>
<H2>School of Printing Management and Sciences</H2>
<IMG SRC = "colorbar.GIF" alt="----------"><P>
The School of Printing Management and Sciences is one of five schools in RIT's <A
HREF="http://www.rit.edu/Academic/Imaging/">College of Imaging Arts and Sciences</A>, which houses
programs in professional photography, art and design, crafts, and imaging science. The school's facilities are
unsurpassed: students learn from more than $33 million worth of up-to-date equipment in 15 laboratories and
43,000 square feet of facilities.<P><HR>
<BLOCKQUOTE><I> "Here we have built a school of printing. Within these strong walls there shall be time
to think; time to learn; time to perfect invaluable skills, and to explore new techniques. Across this threshold
young men shall confidently stride to meet the challenging years ahead, armed with sound training. Stranger,
bid them Godspeed along that sunlit road."</I><B>—Beatrice Warde</BLOCKQUOTE><HR>
<h2>Table of Contents:</h2>
<dl><dd><b>Undergraduate Programs</B>
<dl>
<dd><IMG align=top SRC = "ball.GIF"><a href="printingmanagement.html">Printing Management</a>
<dd><IMG align=top SRC = "ball.GIF"><a href="printingapplied.html">Printing & Applied Computer
Science</a>
<dd><IMG align=top SRC = "ball.GIF"><a href="printingsystems.html">Printing Systems</a>
<dd><IMG align=top SRC = "ball.GIF"><a href="newspaper.html">Newspaper Operations
Management</a></ul><p>
<b>Graduate Programs</B>
<dl>
<dd><IMG align=top SRC = "ball.GIF"><a href="gapublishing.html">Graphic Arts Publishing</a>
<dd><IMG align=top SRC = "ball.GIF"><a href="gasystems.html">Graphic Arts Systems</a>
<dd><IMG align=top SRC = "ball.GIF"><a href="printingtech.html">Printing Technology</a></dl><p>
<b>List of Courses</B>
<dl>
<dd><IMG align=top SRC = "ball.GIF"><a href="ugcourses1.html">Undergraduate Course Listings</a>
<dd><IMG align=top SRC = "ball.GIF"><a href="gcourses.html">Graduate Course Listings</a></dl><p>
<b> Grab Bag</b><IMG SRC = "new.GIF">
<dl>
<dd><IMG align=top SRC = "ball.GIF">A <a href="collection.html">collection</a> of related and current
information<br> pertaining to the School and the Graphic Arts industry.</dl><p>
<b>Related Links</b>
<dl>
<dd><IMG align=top SRC = "ball.GIF">A Page of <a href="links.html">Hotlinks</a> to Graphic Arts related
Web sites.</dl></dl>
<IMG SRC = "rain_lin.GIF" alt="----------">
<center><img src="potgm.GIF" alt=" "> <br>
<b>School of Printing Management and Sciences</b>
<br><br>

<i>Send comments to: <a href="mailto:ejc1753@rit.edu">webmaster</a></i>
<pre>Last update: July 19, 1995</pre></center>
```

Figure 2
Home Page Source Code

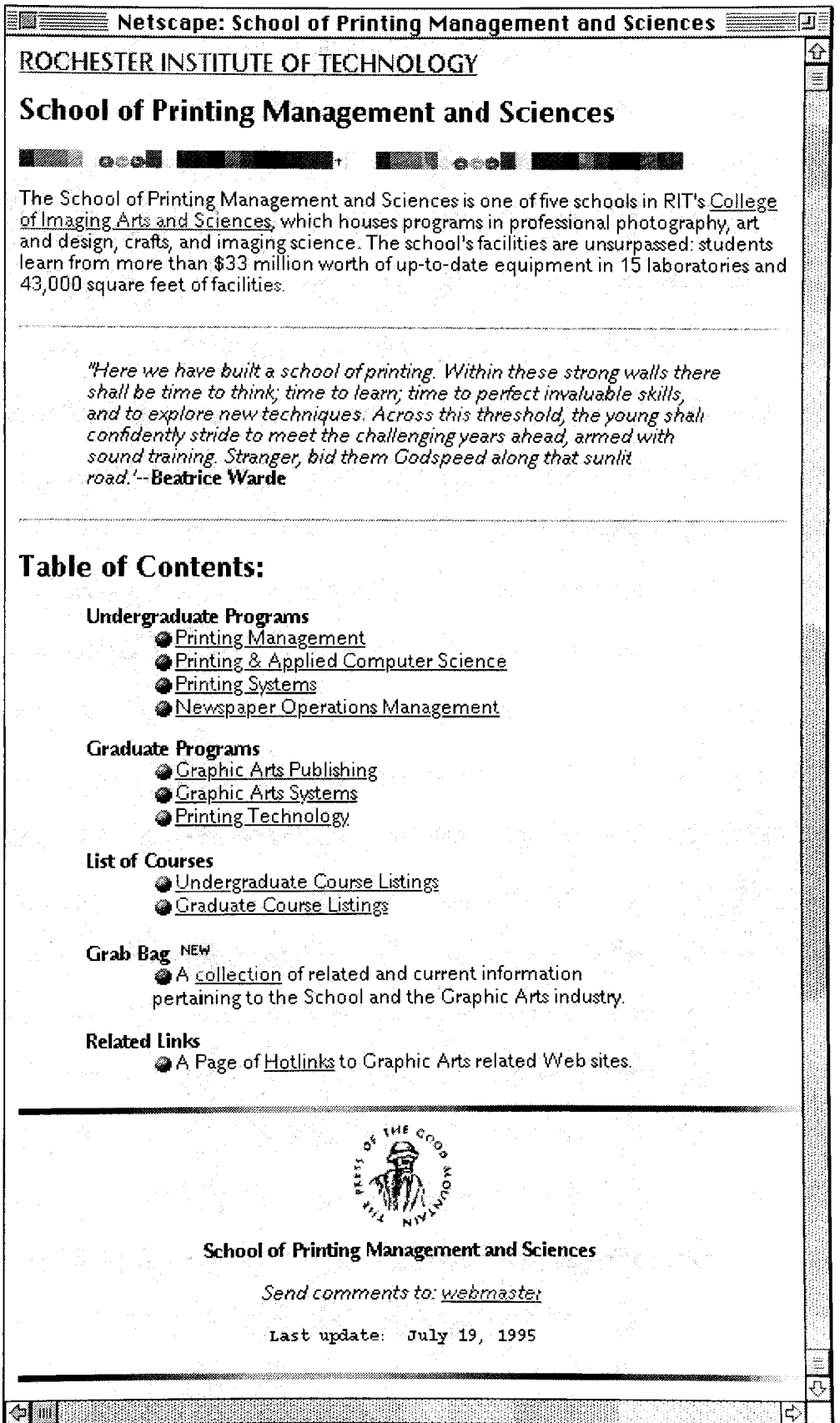


Figure 3
Graphical Representation of Home Page

Style Issues

HTML is based on SGML (the Standard Generalized Markup Language), which is used to describe the general structure of various kinds of documents. It is not a page description language like PostScript, nor is it a language that can be generated by a page layout program. The focus of HTML is the content of the document, not its appearance.

With a few minor exceptions, HTML does not describe the appearance or layout of a document. The designer of a Web page has no control over line length, typeface, point size, or color. These characteristics are specified by the end user according to their Web browser, personal preference, or system limitations.

The obvious disadvantage of this approach is that the designer has limited control over how the document will be viewed. This frustrating situation takes some getting used to. Designers usually come from disciplines where they control the final output, and the thought of someone else controlling the final look of a document is an awkward one. **Figure 4** and **Figure 5** display how the same HTML document will be represented by the Mosaic and Netscape browsers.

Working in a text-only markup language, with little control over the appearance of a document may seem frustratingly archaic compared with today's technology. But for the kind of environment that the Web provides, HTML does have advantages over other forms of document publishing languages. For example, each HTML document is small, so it can be transferred over the Internet as fast as possible. Typeface or format information is not within the document and thus allowing for

faster loading and displaying. Also, HTML documents are device independent and can be displayed on any platform. This is absolutely necessary considering that all types of computer platforms access information off the Internet. All that is needed is a Web browser that can interpret HTML for that platform .

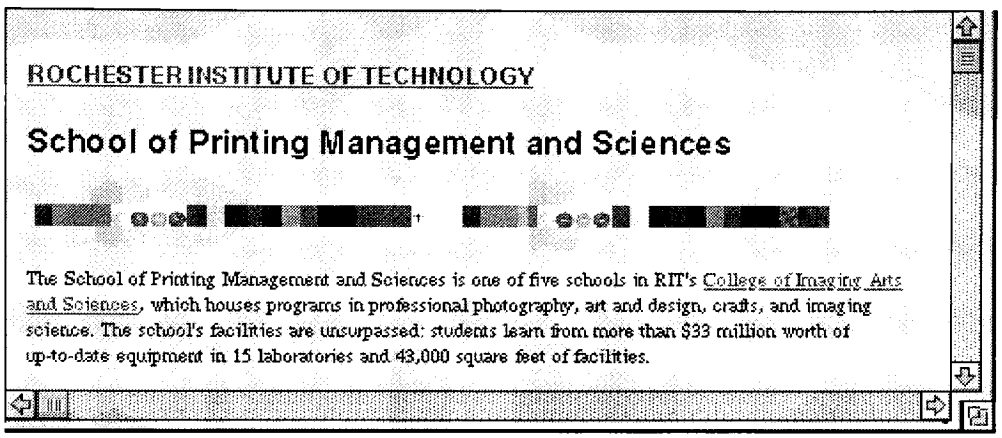


Figure 4
Mosaic representation

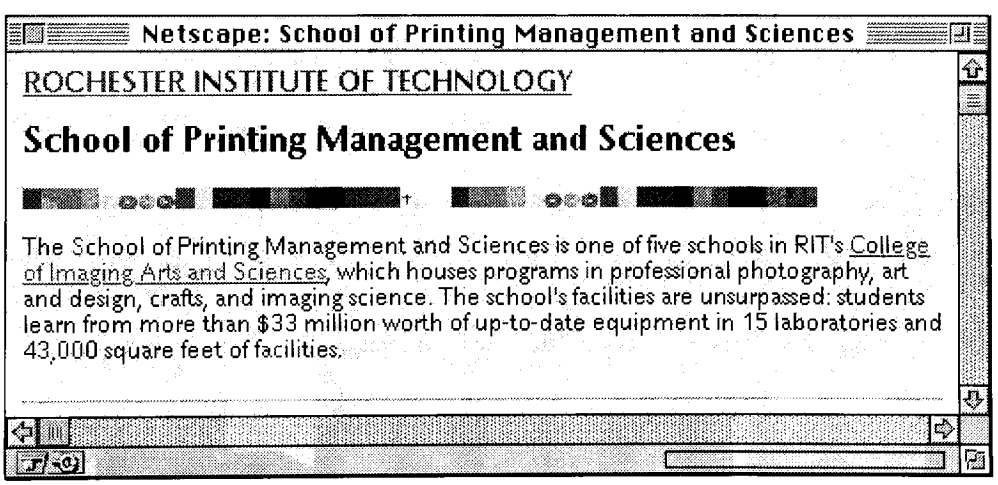


Figure 5
Netscape representation

Implementing Mixed Media Files

Inline versus external images

The IMG (Image) element allows an image to be inserted within an HTML document. This is known as an inline image whose function is to allow graphics to be included within a document and presented with the text. The opposite is an external image which has hypertext links that retrieve and display images in a separate viewing window. Both methods allow a user to display an image, and when used appropriately, can enhance a Web presentation considerably.

Inline Images

Although many different graphics file formats exist, most browsers will only recognize a few. In practice the only format common to all browsers is GIF (Graphic Interchange Format) devised by CompuServe. While this is the most commonly used format on the Web, other graphics file formats have their strengths and some are set to increase in popularity. Browsers that can display images in other formats, primarily JPEG (Joint Photographers Expert Group) are starting to appear. As yet such browsers are not in common use, but are likely to be so within the next year.

GIF was developed to be a device-independent method of storing pictures. GIF allows high-quality, high resolution graphics to be displayed on a variety of graphics hardware and is intended as an exchange and display mechanism for graphic images.

GIF is reasonably well matched to inexpensive computer displays, since it only stores 8 bits/pixel and most PCs can not display more than 256 distinct colors at once. GIF does well on images with only a few distinct colors, such as line drawings and simple cartoons. A GIF picture file has an extension ".gif."

Transparent GIFs are useful because they appear to blend in smoothly with the user's display, even if the user has set a background color that differs from that the developer expected. This is done by assigning one color to be transparent—if the web browser supports transparency, that color will be replaced by the browser's background color, whatever it may be.

There are two methods for presenting images. Interlaced GIFs appear first with poor resolution and then improve in resolution until the entire image has arrived, as opposed to images arriving linearly from the top row to the bottom row. The interlaced approachis great to get a quick idea of what the entire image will look like while waiting for the rest. Some browsers do not support progressive display as the image is downloaded, but non-progressive-display web browsers will still display interlaced GIFs once they have arrived in their entirety.

JPEG (Joint Photographers Expert Group) is a standardized image compression mechanism. JPEG is the original name of the committee that wrote the standard. JPEG is designed for compressing either full-color or gray-scale images of natural, real-world scenes. It works well on photographs, naturalistic artwork, and similar material; but poorly on lettering, simple cartoons, or line drawings.

JPEG stores full color information: 24 bits/pixel (16 million colors). Therefore, with full-color hardware, JPEG images look much better than GIFs on most platforms. JPEG files are much smaller than GIFs, therefore, they are superior to GIF in terms of disk space saving and transmission time. A JPEG file has an extension ".jpg."

The JPEG standard is an excellent standard for most realistic images (photos for example, but not line drawings or logos). It uses a powerful, though nominally "lossy", compression method. JPEG is best suited for truecolor original images; avoid using it on images that have already been forced into a 256-color palette.

Using JPEG for a photographic image for example can produce 10:1 savings compared to GIF, as well as permitting much better display quality on truecolor-capable displays. Netscape handles inline JPEG; most older browsers need to use an external JPEG viewer.

External Images

The only inline image file format that all graphic browsers can view is GIF. Linking to external images gives more flexibility in what image file format can be used. **Table 1** shows the most popular external image file formats with their corresponding extensions. It is essential for external images to have the correct file extension so the Web browsing software knows how to interpret it or be able to launch the correct helper application.

Format	Extension
GIF	.gif
JPEG	.jpg or .jpeg
PICT	.PICT
XBM	.xbm

Table 1
External Image formats and extensions

A common practice with Web pages is to provide a very small GIF image (a "thumbnail") inline on the page itself, and then link that image to its larger counterpart. This has two major advantages over including the entire image inline: It keeps the size of the Web page small, so that page can be downloaded quickly and it lets the reader get a feel for the image and decide whether it it suitable to download the whole image. This flexibility is an advantage when providing information to a large audience with a myriad of different systems and requirements.

Sound

Sound data is a digital representation of an analog signal, which is typically represented as a continuous waveform. When sound is digitally recorded, samples of wave form are captured at fixed intervals. The more samples that are taken, the more information stored for each sample, the higher quality of the sound. Sample rate measured in kilohertz (KHz) describes the quality of the sound. Common sample rate are 11 KHz, 22 KHz and 44 KHz. Sample size, usually in 8-bits or 16 bits, is the amount of information stored for each sample. For example, audio CD has 44 KHz and 16 bits sample size, which is very high quality sound.

8000 sample 8 bit .au files are currently the best choice for putting sound files on the Web. The reasons for this choice are: this format can be supported by most of the sound-capable machines on the Internet. Secondly, this format produces a fairly small file that requires little bandwidth to transfer. AU files are of only barely acceptable quality, as the 8-bit sampling causes them to sound comparable to transmission over a telephone.

There are other sound formats supported by Web browsers. It is important to note that most sound formats are specific to a certain platform. **Table 2** provides a listing for the sound formats found on the Internet.

Format	**Extension**
AU/μ-law	.au
AIFF	.aiff
WAV	.wav
MPEG Audio	.mp2

Table 2
Sound formats and extensions

Video

The four variables which affect the size of a video file are the length of the clip, the size of the frame, the color depth (256 colors, thousands, millions, etc.), and the amount of compression applied. Manipulation of all these settings determines what combination produces acceptable quality files with the minimum size.

Like other files, movie files can be identified by their file extensions. There are only few movie file formats that can be viewed from the Internet, which are international standard file formats for multimedia. They are MPEG and Quicktime.

Format	Extension
MPEG	.mpg
Quicktime	.mov & .qt

Table 3
Video formats and extensions

MPEG is a very popular movie file format for PCs and stands for Moving Pictures Expert Group. The members of this group come from more than 70 companies and institutions worldwide including SONY, Philip, Matsushita and Apple. They meet under the International Standard Organization (ISO) to generate digital video standards for Compact Discs, Cable TV, Direct Satellite Broadcast and high-definition television.

Quicktime is an ISO standard for digital media. It was originally created by Apple Computer Inc. and used the in Macintosh. It brings audio, animation, video, and interactive capabilities to personal computers and consumer devices. QuickTime movies are real movies. This standard is more mature than the MPEG standard. In December, 1993, Apple announced that it had begun demonstrating technology that will make future television and multimedia devices more compelling, interactive, and useful for people. Specifically, Apple demonstrated the integration of MPEG technology into applications using QuickTime technology. QuickTime for Windows is available for customers who use Microsoft's Windows/DOS operating system.

QuickTime movies have file extension ".qt" and ".mov." .Mov files can be played on both Macintosh and PCs, while .qt extensions can only be viewed on a Macintosh.

There are other multimedia file formats, for example, .avi is a video format for Microsoft Windows, .awa/.awm are Gold Disk Animation, they are not broadly used on the Internet. However, there are some file converters that convert .avi files to .mov files.

Server Issues

The chief hardware component of a Web site is the host computer, which houses the site's content. The software program that runs on the host is called a Web server. It interprets incoming requests and returns the appropriate documents. Server programs are available for most major operating systems, including Unix, VMS, VM, Macintosh System 7, Windows 3.1, Windows NT.

The major considerations in choosing a World Wide Web host are platform stability, performance, and amount of RAM (random access memory). Platform stability is most important. If the operating system routinely crashes under normal loads, the Web site often will be inaccessible.

In general, Unix computers have the most stable and robust operating systems, but their stability can be outweighed by their cost and technical requirements. Macintosh and Windows machines are acceptable for handling light loads (with peaks of several hundred hits per hour), but are not recommended for higher loads

Most of today's business hardware—including 486-based PCs, 68040-based Macs, and nearly every available Unix workstation—can handle the processor and I/O (input/output) demands of all but the most heavily trafficked Web servers.

However, the host computer's physical RAM can cause a bottleneck as the usage load increases.

A basic WWW platform should have at least 8MB RAM, and high-load computers should include 64MB RAM. To serve heavy loads, the host computer should be a dedicated WWW machine—one that is used exclusively for the Web site. Under low loads a host can be used for tasks in addition to its Web server duties.

HTTP

HTTP (Hypertext Transport Protocol) is the primary protocol used to distribute information within the World Wide Web. It is a relatively simple, highly flexible protocol used to deliver information across the Internet. HTTP defines a simple transaction, consisting of the following four parts, to deliver requested information from a server to a client:

The client must first establish a connection to the server.

The client then issues a request to the server specifying a particular document to be retrieved.

The server sends a response containing a status code and the text of the document if it is available.

Either the client or the server then disconnects.

One main goal of HTTP was to provide a simple algorithm that would enable fast response times. To achieve this goal, HTTP was defined as being a "stateless" protocol—one that does not retain any information about a connection from request to request.

HTTP is also a "connectionless" protocol—limited to one request per connection. Unlike other protocols, such as FTP, the connection between server and client is broken after each request is made. This means that every time a client wants to fetch a document, it establishes a new connection to the HTTP server.
This is one of the main reasons why it takes so long to load HTML pages with many inline graphics. For each graphic, a separate connection is established and information is requested. While establishing a connection is not generally time consuming, it can seriously affect performance for distant or heavily loaded sites.

Some clients, like Mosaic, wait until a connection is closed before they open their next connection. However, many newer clients, such as Netscape, open multiple connections and receive documents in parallel. Unless bandwidth is the bottle-neck in retrieving documents, this behavior results in a significant time savings when accessing sites loaded with inline images.

Chapter 7: Summary and Conclusions

Data Size

The amount of data that has to be transported for image or video files is the most important concern for WWW publishing. This is especially important because of the large number of users on 56-Kbps or slower lines. For them, it can take minutes to receive even medium-sized images, sounds, or video files.

Browsers are beginning to cater more to users on slow lines by providing extra functionality. Most browsers let the user delay or completely turn off image-loading functions. Several second-generation browsers request images simultaneously, or in parallel, which significantly speeds the retrieval process. However, while these features help, content providers must carefully watch the size of the images they provide.

There are a number of techniques for reducing the data size of images. Although the most obvious way may seem to reduce the images total physical size, this does not necessarily yield substantial data-size reductions. An important issue to keep in mind when generating images is that different data formats compress images differently. For a particular image, the JPEG format may be significantly smaller than the GIF format, or vice versa.

Although it may not seem apparent, reducing the number of colors in an image can produce substantial savings in data size. By halving the total number of colors, the file size can be reduced by 10 percent to 30 percent. Luckily, the change in the number of colors usually is barely detectable to users. Because most browsers already limit the number of colors seen by users (typically 25 to 50 col-

ors), reducing the number of colors makes an image more in accord with the browser display.

The problem of data size can become significantly worse when dealing with audio and video files. A ten second low-quality audio clip can take more than a minute to download over a 14.4-Kbps line. The size of higher-quality audio clips—in stereo or at higher-sampling rates—is at least double that of low-quality clips.

Depending on their quality and duration, video clips generally range from hundreds of kilobytes to 50MB or more. Even a few seconds of low-quality video—without audio—can cause significant waiting on a 56-Kbps line. For video as small as thirty seconds, may take minutes even over a full T1 (1.54 Mb/sec.) connection. For this reason, audio and video files should be used sparingly and should be marked carefully and appropriately as large files.

The best means of reducing the data size for audio or video files is to reduce their quality. For example, changing from stereo to mono sound can yield a data-size reduction as large as fifty percent. Lowering the frame rate or resolution in a video file or the sampling frequency in an audio file can substantially reduce the file data size.

Physical Size

Another important consideration is an images physical size (height and width). Although it is directly related to data size, it is important to keep the physical size within reasonable limits so it looks acceptable on most viewers.

To fit completely within the windows of most Web browsers, images should not be wider than six inches. All images to be served on the Web should fit within this limit. Most image-editing programs can perform these functions.

One popular and useful way to reduce the size of the graphics on a page is to use thumbnail images as links to the main images. Thumbnails are small versions (approximately one inch square) of images. Besides being small (2K to 5K), they can save time by letting users decide whether or not to retrieve the full image.

Similarly, it is important to use reasonable-length video and audio clips (less than one minute). On some systems, it is difficult to stop a video or audio file from playing once it is started. Therefore, in addition to keeping the data size of the file small, use shorter clips to give your users greater control over what they have to view.

Size and Format Indicators

It is considered good etiquette to label all large files (larger than about 100K). This applies to graphics, audio, video, and other large file formats. Because some clients can only accept certain data formats, it also is helpful to provide descriptors or icons that identify a file as image, audio, or video and to indicate its data format and size. This is especially true of very large files. It is extremely frustrating for users to download a 5MB video file only to discover that it is in a format not supported by their browsers.

Text-Only Users

Although a large number of Web users run graphical interfaces, text-based Web clients such as Lynx continue to be popular. Many Web users also use their graphical browsers in text-only mode to speed document retrieval. These users are often ignored, reducing the value of a Web site.

The most frequent mistake made by content providers is not using functional ALT tags for inline images. An ALT (alternative) tag is embedded within an HTML document, so non-graphical browsers can display something other than the default `[IMAGE]` label. This means that a site viewed with Lynx will contain a large number of holes with `[IMAGE]` labels. Perhaps even worse are the ALT tags that describe an image. Reading "a picture of our corporate logo" instead of viewing the logo does little for users. Browsers such as Lynx will insert the text directly into the document, which can confuse readers.

ALT tags should be functional, not descriptive. The ALT tag for a bullet icon should be "o" not "Bullet." A graphic used for a horizontal line should be a series of dashes ("-"), not the phrase "horizontal line." Always keep users with text-based clients in mind when writing HTML. This requires little effort and will greatly improve the utility and appearance of the site for those users. **Figure 6** illustrates this concept by utilizing a series of dashes to create a horizontal separation.

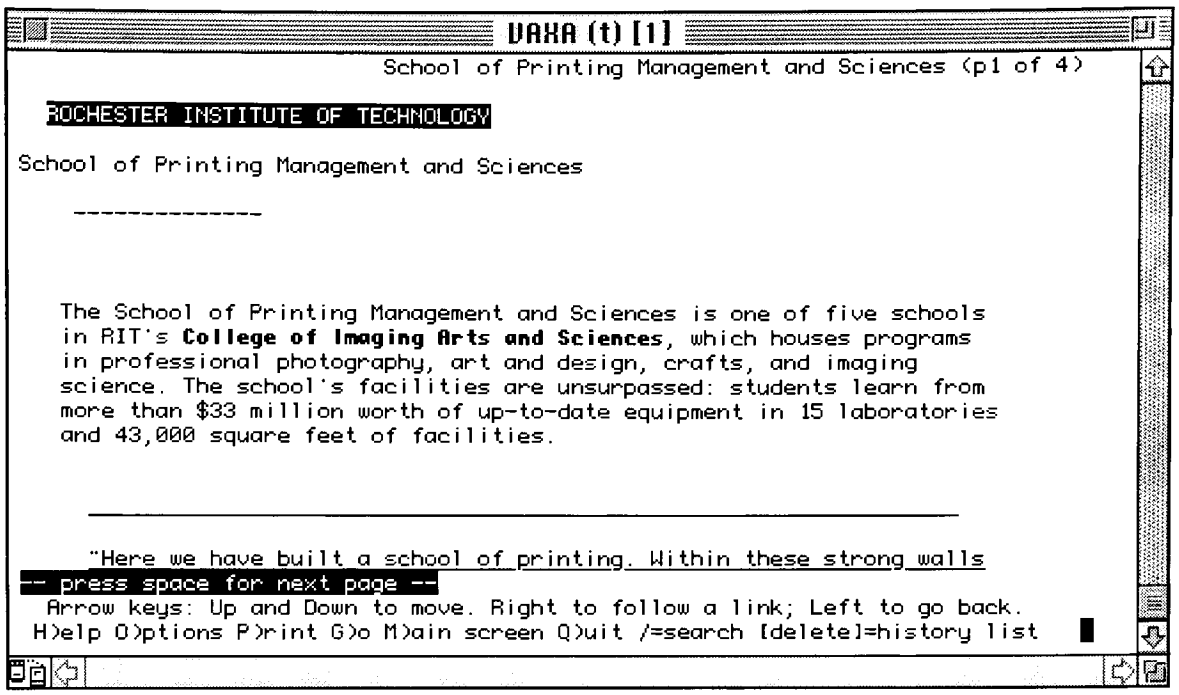


Figure 6
Text-only representation

The goal of this thesis project was to test and demonstrate the World Wide Web as a publishing vehicle by creating a World Wide Web presence for the School of Printing Management and Sciences. In order to reach this goal, a full understanding of the Hypertext Markup Language was realized. Once this is accomplished, issues regarding integration of mixed-media elements within an HTML document were investigated. Once a prototype of the HTML document was complete, the mixed-media elements were tested and evaluated for proper integration and contextual cohesiveness. Many issues regarding implementation of mixed-media elements, such as file size and file format were addressed upon testing.

One of the additional goals of this project was a comprehensive description of the methodology for creating and maintaining a World Wide Web publishing pres-

ence. The methodology addresses issues such as: navigational software, structuring HTML documents, hypertext linking, HTML style issues and limitations, effective integration of mixed-media elements, inline and external image issues, testing documents, advertising documents, strategies for determining proper file sizes and formats of mixed-media elements, integrating supplemental programs, World Wide Web Server issues, installing HTML and mixed-media files onto a World Wide Web Server, etc.

The Web's format and the opportunity to reach millions of potential customers is creating new types of publishing ventures in true "gold-rush" fashion. The Web is being touted as the fourth medium, and some suggest it will have as great an impact on society as print, radio and television. The growth of the Web is explosive and will assuredly continue to grow.

Upon completion of this study, the author remains skeptical whether the World Wide Web is the medium of the future. It has, however, created a trend which will forever reshape the publishing world and the way information seekers receive their data. Publishing will change from a commodity based market where prices are based upon cost, and shift to a service market where prices are based upon the value of the information. Each reader requiring selected information tailored to their specific choice and paying for what they select—Not paying for an entire magazine or newspaper in order to read only one article. The future of information dissemination is electronic, interactive and selective. Whether the delivery mechanism will be the World Wide Web remains to be seen.

Recommended areas for further study

Currently, the Web is the most advanced information system deployed on the Internet, and is equipped to embrace many future advances in technology, including new networks, protocols, and data formats. The most prominent shortcoming of the Web is its inability to securely transfer private data. Further study into data encryption over computer networks is necessary before the Web can realize its full potential. Once secure transactions can take place, the Internet will become a mass means of both information and commerce.

Utilizing the World Wide Web as a publishing vehicle is at the mercy of available bandwidth. To take full advantage of the Web's hypermedia capabilities requires a considerable amount of bandwidth. Continued research into data compression can alleviate a host of data transfer issues currently limiting the functionality of the Web. With data traveling at faster speeds, designers of Web pages have more flexibility in the information they offer. It is the author's opinion that increased bandwidth results in more dynamic Web documents.

Bibliography

Bibliography

Branwyn, G. "Mosaic Quick Tour." Chapel Hill: Ventana Press.

Cutler, M. "Setting Up Your Own Web Site." *Web Watch.* April 95.

Dies, Jeri H. "We Got Culture!" *Internet World.* v 6 n 4 April 1995.

Editor. "An Express Lane for the Infobahn." *Business Week.* 24 October 1994.

Editor. "And Leading the News in IBM's Cyberzine..." *Business Week.* 14 November 1994.

Editor. "Electronic storefronts on the Internet." *Fortune* 28 November 1994.

Editor. "Electronic Document Delivery." *Seybold Special Report* 10 October 1994.

Editor. "How to Do Business on the Web." *PC World.* v 13 n 3 March 1995.

Editor. "Internet and Online Publishing." *Seybold Special Report.* 10 October 1994.

Editor. "Internet Publishing Tools Proliferate." *Byte.* 1 March 1995.

Editor. "Online." *Information Week.* 16 January 1995.

Editor. "Publishing on the Net Sports Industry Resurgence." *Seybold Special Report* 10 October 1994.

Editor. "The Internet: How It Will Change the Way You do Business." *Business Week.* 14 November 1994.

Editor. "Tools for Internet Publishing. "*Seybold Special Report.* 10 October 1994.

Editor. "Web Publishing: Pieces Falling into Place." *Seybold Report on Desktop Publishing.* 10 October 1994.

Engst, Adam. "Internet Starter Kit." Indianapolis: Hayden Books, 1994.
Ford, Andrew. "Spinning the Web." London: Thomson Publishing, 1995.

Haiman, Steven. "The World Wide Wonder." *Color Publishing.* March/April 1995.

Hayes, Brian. "The World Wide Web." *American Scientist.* v. 82 September/October 1994.

Lemay, Laura. "Teach Yourself Web Publishing with HTML in a Week." Indianapolis: Sams Publishing.

Lewis, Peter, H. "Cruising the Web with a Browser." *New York Times.* 7 February 1995.

Lewis, Peter H. "Making it Easier to Enter the Web." *New York Times.* 24 January 1995.

Negroponte, Nicholas P. "Homeless' info.hwy.net." *New York Times.* 11 February 1995.

Powell, James. "Spinning The World-Wide Web: An HTML Primer." *Database.* v 18 n 1 February 1995.

Press, Larry. "Commercialization of the Internet." *Communications of the ACM.* v. 37 November 1994.

Savetz, Kevin. "The Medium Is The Matrix." *Internet World.* v 6 n 4 April 1995.

Semich, J. William. "The World Wide Web: Internet Boomtown?" *Datamation.* 15 January 1995.

Sheldrick, Dennis." Start Becoming a Webmeister." *Datamation.* 1 March 1995.

Stefanoe, S. "Multimedia Online." *New Media.* November 1994.

Vacca, John R. "Mosaic: Beyond Net Surfing." *Byte.* 20 January 1995.

Vaughan-Nichols, Steven J. " The Web Means Business." *Byte.* 19 November 1994.

Vetter, Ronald J. "Mosaic and the World-Wide Web." *Computer.* v. 27 October 1994.

Weiss, Aaron. "Hop, Skip, and Jump." *Internet World.* April 95.

Wiggins, R. "Webolution," *Internet World.* April 1995: 33-38).

Appendices

Appendix A

Appendix A

HTML Tag Reference

HTML is composed of a set of elements that define a document and guide its display. An HTML element may include a name, some attributes and some text or hypertext, and will appear in an HTML document as:

<tag_name> text **</tag_name>**

<tag_name attribute_name=argument> text **</tag_name>**, or just

<tag_name>

For example:

<title> My Useful Document **</title>**

and

<pre width=60> A lot of text here. **</pre>**

An HTML document is composed of a single element:

<html> . . . **</html>**

that is, in turn, composed of head and body elements:

<head> . . . **</head>**

and

<body> . . . **</body>**

To allow older HTML documents to remain readable, **<html>, <head>**, and **<body>** are actually optional within HTML documents.

Elements restricted to the head element

<isindex>
Specify index file
<title> . . . </title>
Specify document title
<nextid>
Set a variable value. Attribute: variable name
<link>
Specify relationships to other documents. Attributes: same as Anchor below.
<base>
Specify the name of the file in which the current document is stored. This is useful when link references within the document do not include full pathnames (i.e., are partially qualified)

The following sections describe elements that can be used in the body of the document.

TEXT ELEMENTS

<p>
The end of a paragraph that will be formatted before it is displayed on the screen.
<pre> . . . </pre>
Identifies text that has already been formatted (preformatted) by some other system and must be displayed as is. Preformatted text may include embedded tags, but not all tag types are permitted. Attribute: width
<listing> . . . </listing>
Example computer listing; embedded tags will be ignored, but embedded tabs will work
<plaintext>
<blockquote> . . . </blockquote>
Include a section of text quoted from some other source.

HYPERLINKS OR ANCHORS

** . . . **
Define a target location in a document
** . . . **
Link to a location in the same file
** . . . **
Link to another file

** . . . **
Link to a target location in another file

HEADERS

<h1> . . . </h1>
Most prominent header
<h2> . . . </h2>
<h3> . . . </h2>
<h4> . . . </h4>
<h5> . . . </h5>
<h6> . . . </h6>
Least prominent header

LOGICAL STYLES

** . . . **
Emphasis
** . . . **
Stronger emphasis
<code> . . . </code>
Display an HTML directive
<samp> . . . </samp>
Include sample output
<kbd> . . . </kbd>
Display a keyboard key
<var> . . . </var>
Define a variable
<dfn> . . . </dfn>
Display a definition
<cite> . . . </cite>
Display a citation

PHYSICAL STYLES

** . . . **
Bold font
** . . . **
Italics
<u> . . . </u>
Underline
<tt> . . . </tt>
Typewriter font

Definition List/Glossary:
<dl>
<dt> First term to be defined
<dd> Definition of first term
<dt> Next term to be defined
<dd> Next definition
</dl>

The **<dl>** attribute compact can be used to generate a definition list requiring less space.

Present an Unordered List:
<ul>
<li> First item in the list
<li> Next item in the list
**
**

Present an Ordered List:
<ol>
<li> First item in the list
<li> Next item in the list
**
**

Present an interactive menu:
<menu>
<li> First item in the menu
<li> Next item
</menu>

Present a directory list of items:
<dir>
<li> First item in the list
<li> Second item in the list
<li> Next item in the list
</dir>

Items should be less than 20 characters long.

ENTITIES

&keyword;

Display a particular character identified by a special keyword. For example the entity & specifies the ampersand (&), and the entity < specifies the less than (<) character. Note that the semicolon following the keyword is required.

&#ascii_equivalent;

Use a character literally. Again note that the semicolon following the ASCII numeric value is required.

<!-- text >

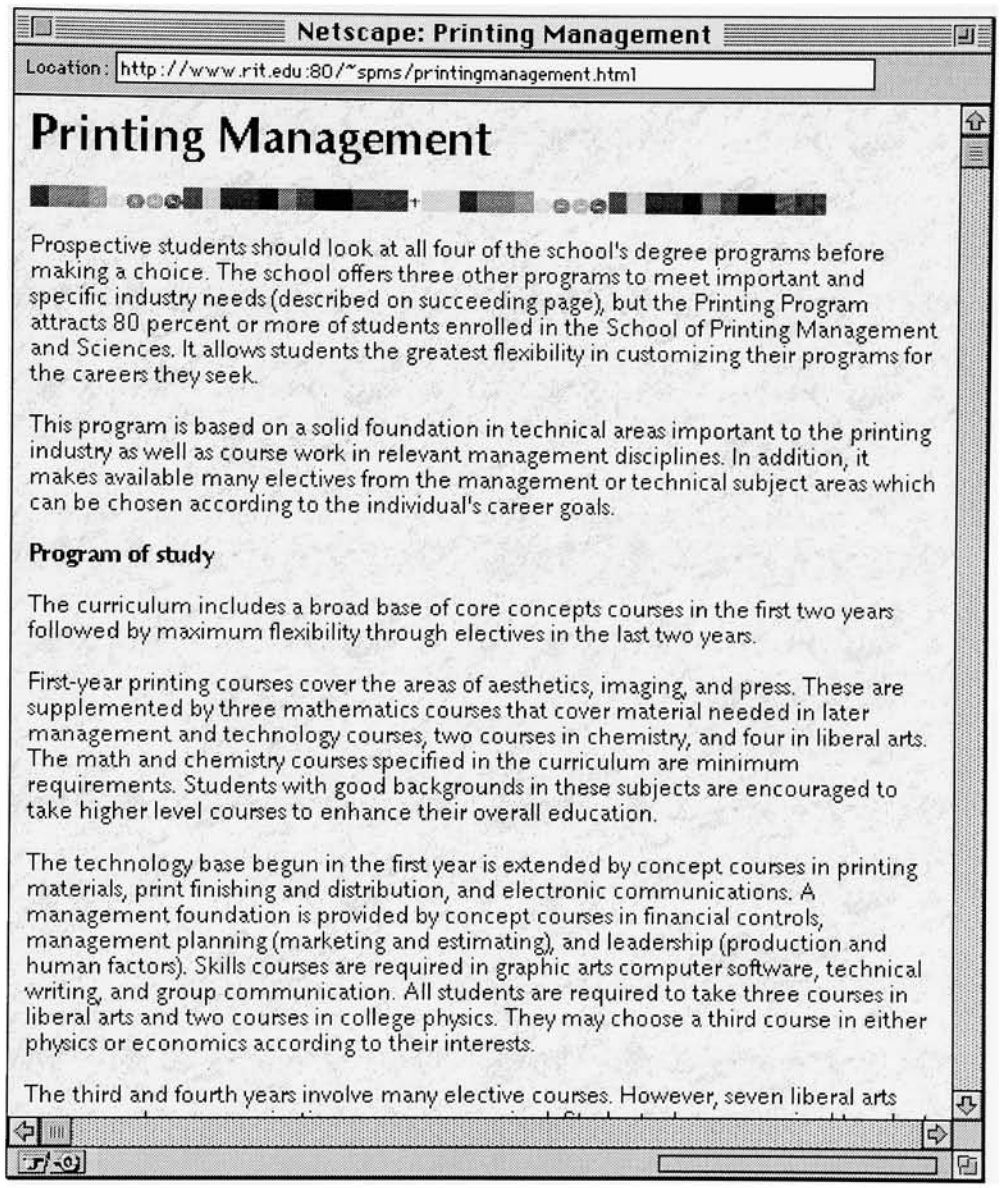
Place a comment in the HTML source

<address> . . . **</address>**

Present address information

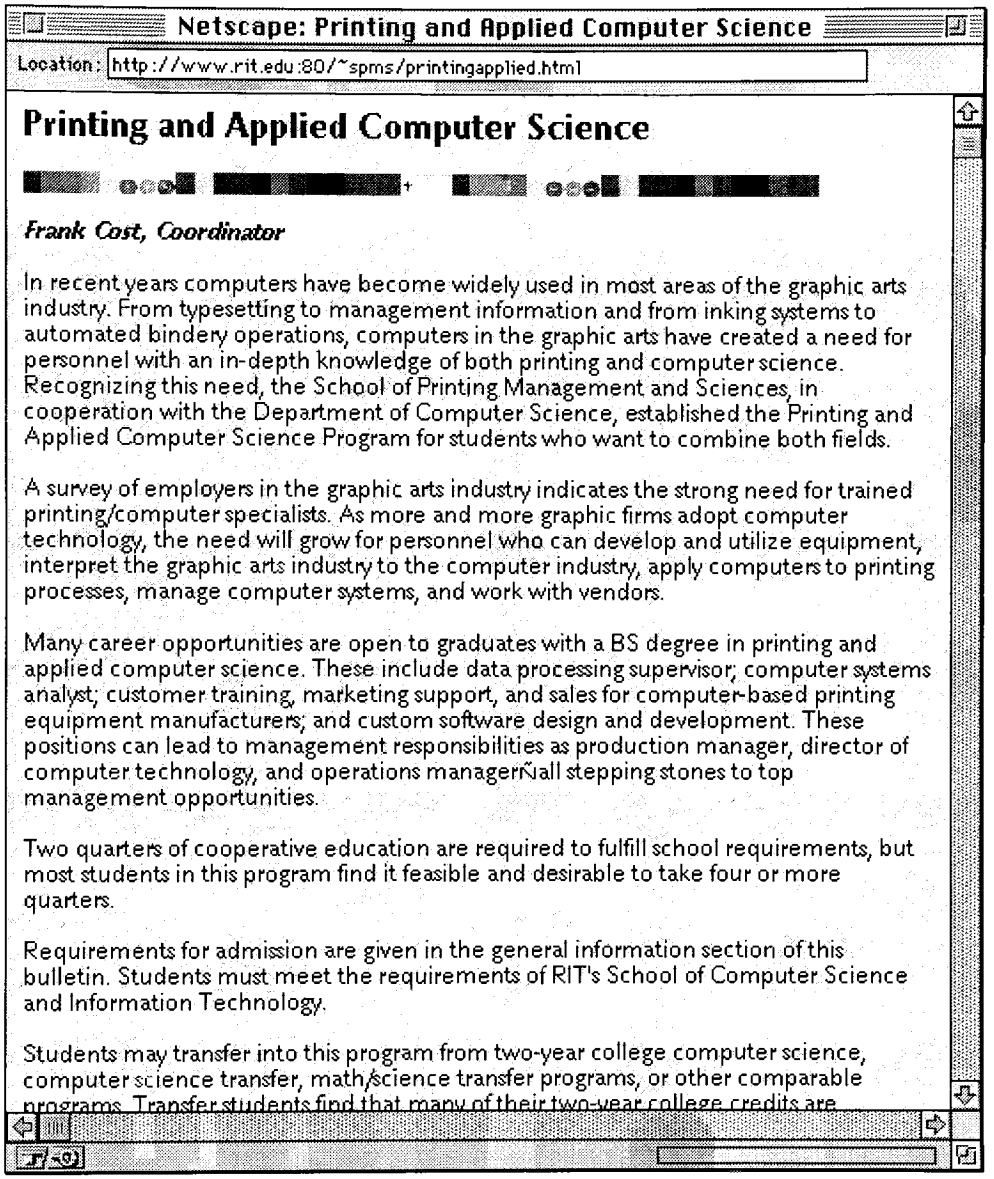
Appendix B

Appendix B



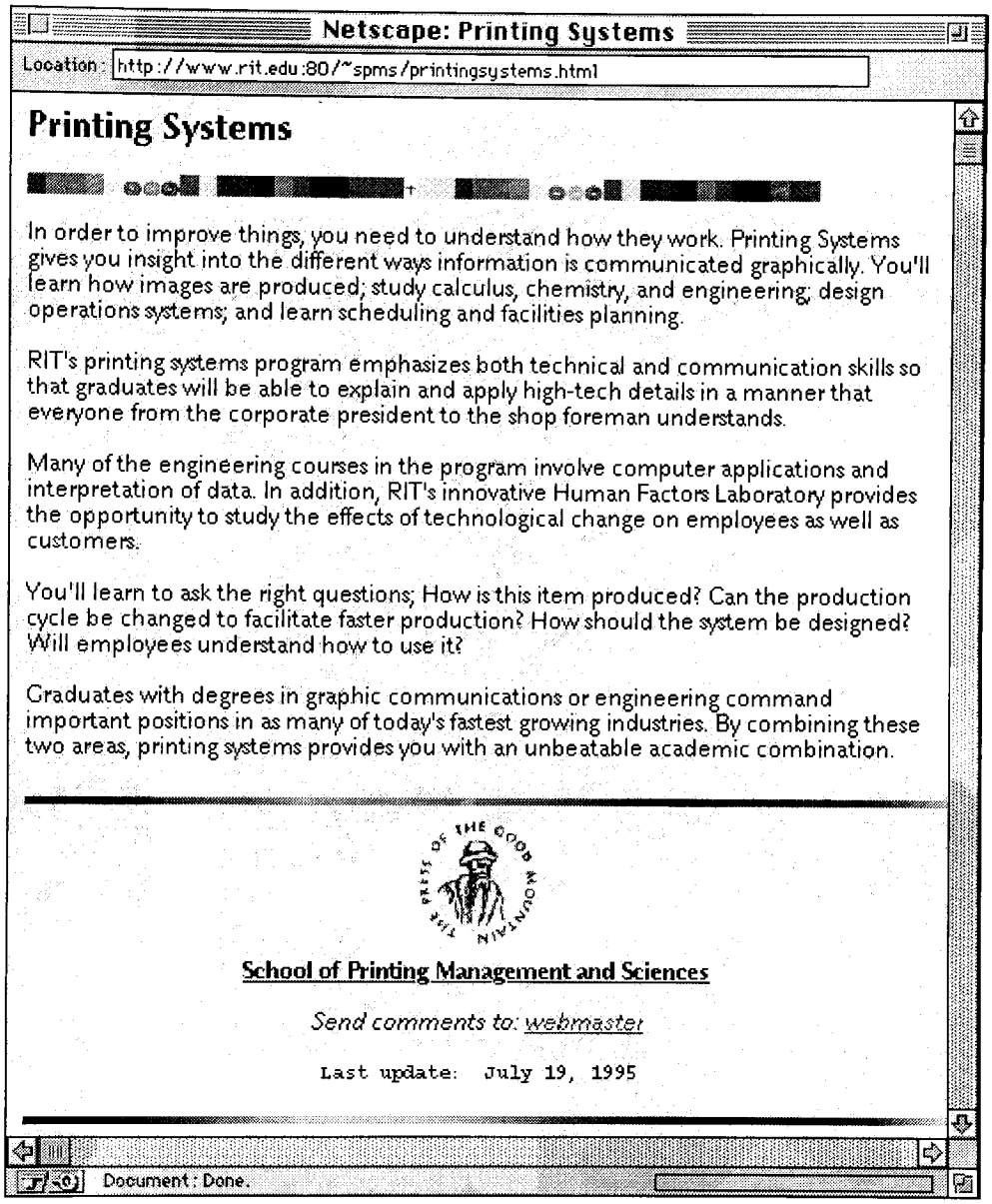
Source Code

```
<html>
<head><TITLE>Printing Management</TITLE></head>
<body><BODY BACKGROUND="parchmnt.gif">
<h1>Printing Management</h1>
<IMG SRC = "colorbar.GIF"><p>

Prospective students should look at all four of the school's degree programs before making a
choice. The school offers three other programs to meet important and specific industry needs
(described on succeeding page), but the Printing Program attracts 80 percent or more of students
enrolled in the School of Printing Management and Sciences. . . .
```



Source Code

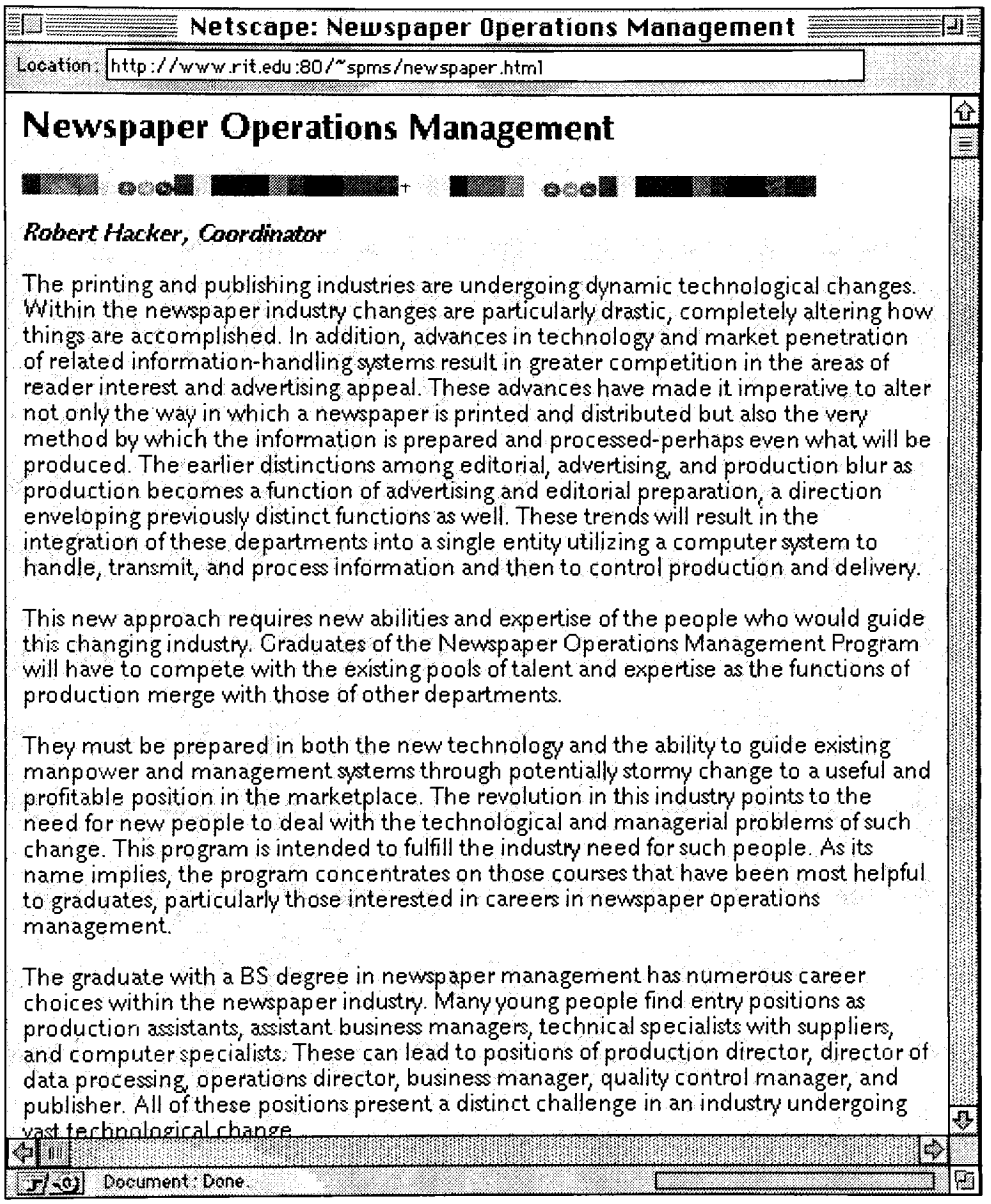
```
<html>
<head><TITLE>Printing and Applied Computer Science</TITLE></head>
<body><BODY BACKGROUND="parchmnt.gif">
<h2>Printing and Applied Computer Science</h2>
<IMG SRC = "colorbar.GIF"><p>
<h4><i>Frank Cost, Coordinator</i></h4>

In recent years computers have become widely used in most areas of the graphic arts industry.
From typesetting to management information and from inking systems to automated bindery
operations, computers in the graphic arts have created a need for personnel with an in-depth . .
```



Source Code

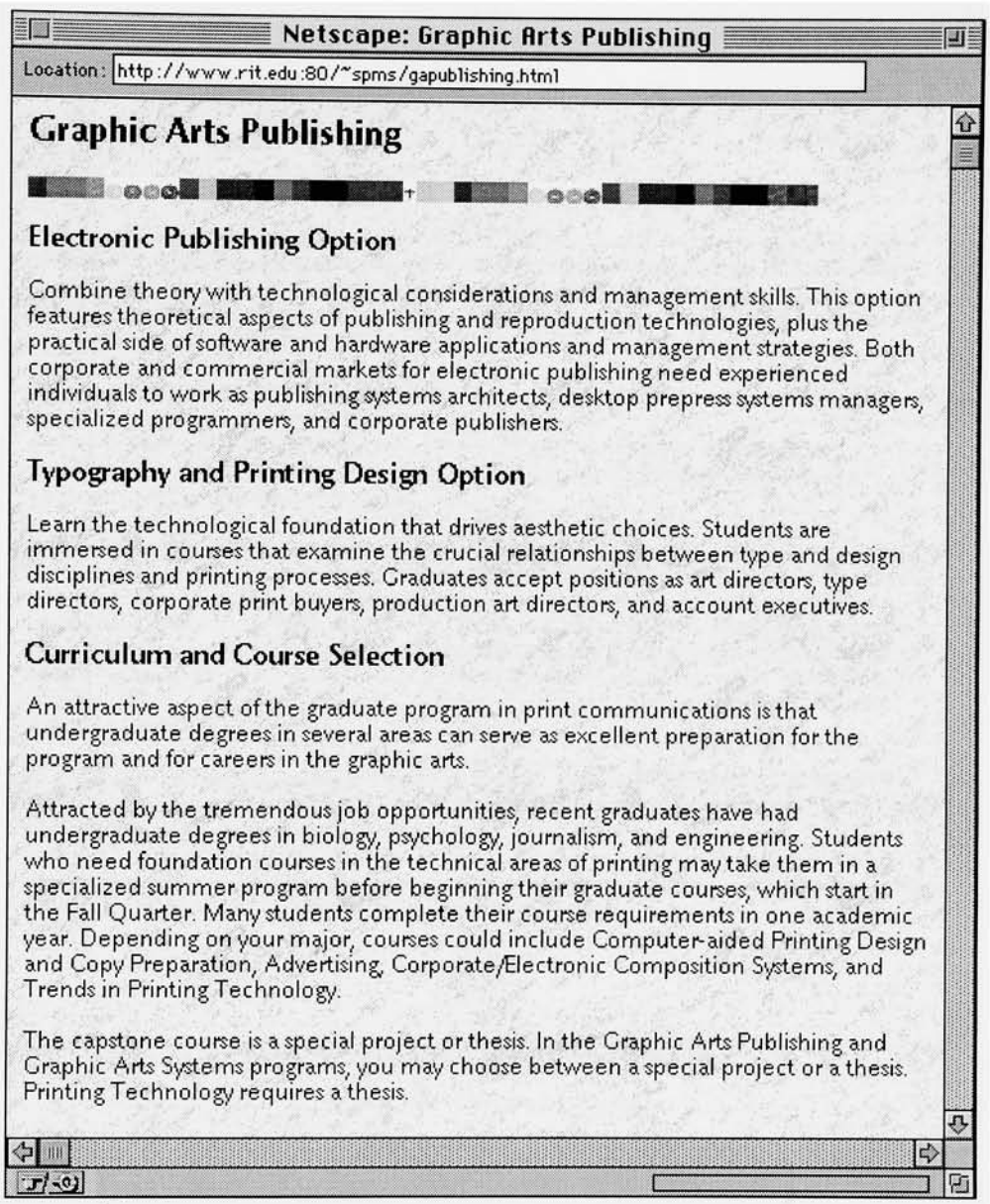
```
<html>
<head><TITLE>Printing Systems</TITLE></head>
<body><BODY BACKGROUND="parchmnt.gif">
<h2>Printing Systems</h2>
<IMG SRC = "colorbar.GIF"><p>
```

In order to improve things, you need to understand how they work. Printing Systems gives you insight into the different ways information is communicated graphically. You'll learn how images are produced; study calculus, chemistry, and engineering; design operations systems; and learn scheduling and facilities planning.<p> . . .



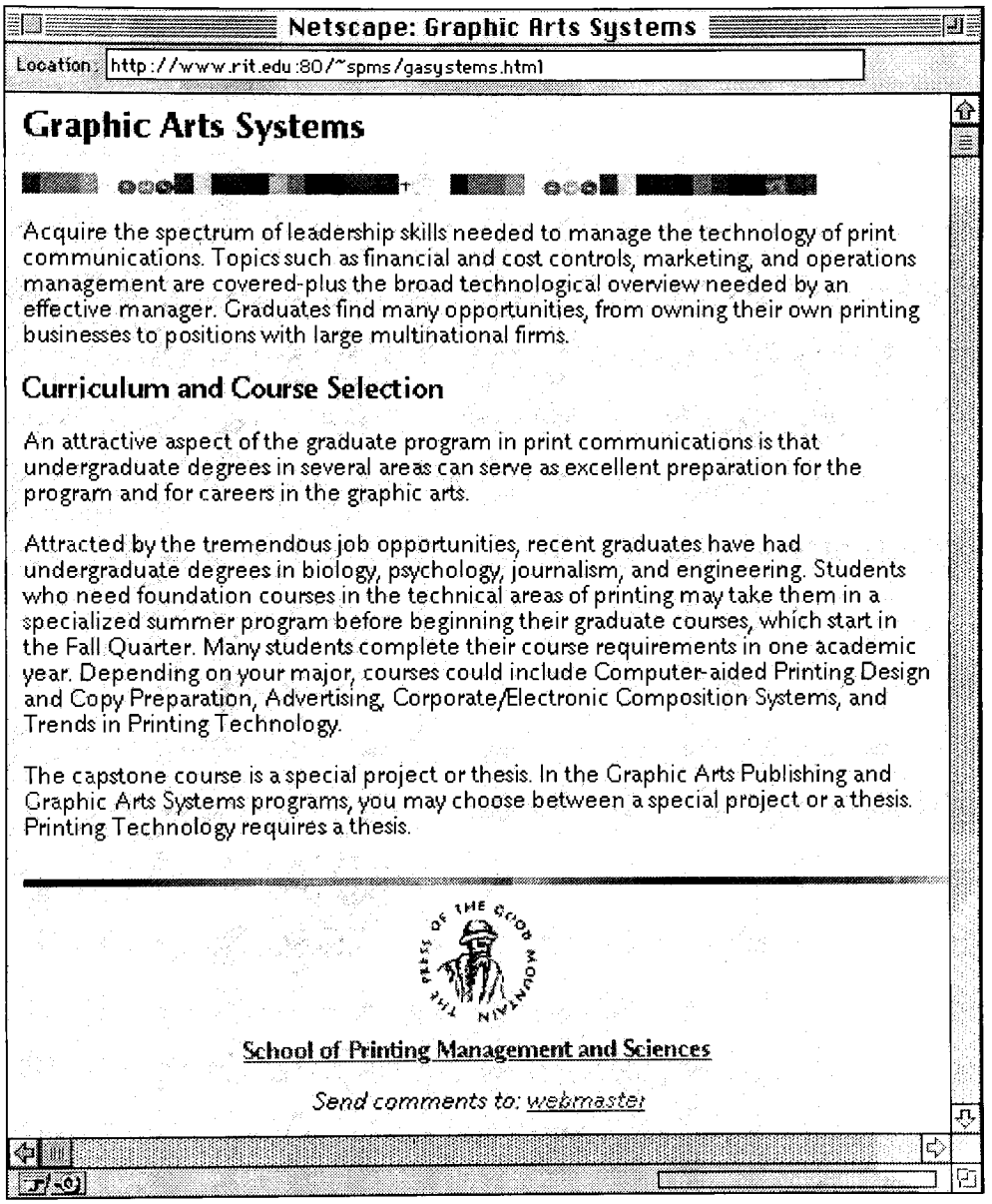
Source Code

```
<html>
<head><TITLE>Newspaper Operations Management</TITLE></head>
<body><BODY BACKGROUND="parchmnt.gif">
<h2>Newspaper Operations Management</h2>
<IMG SRC = "colorbar.GIF"><p>
<h4><i>Robert Hacker, Coordinator</i></h4>

The printing and publishing industries are undergoing dynamic technological changes. Within
the newspaper industry changes are particularly drastic, completely altering how things are
accomplished. In addition, advances in technology and market penetration of related . . .
```



Source Code

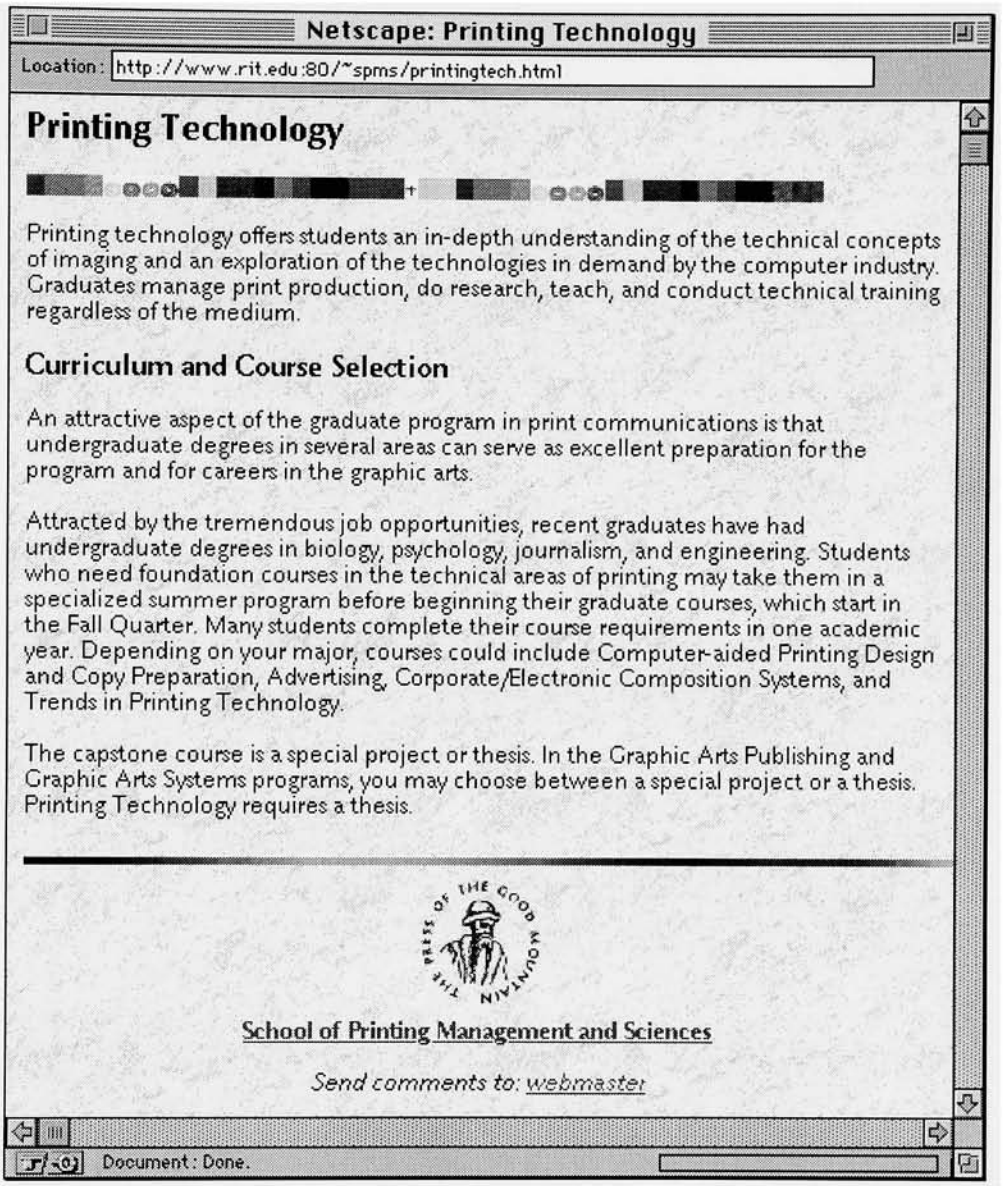
```
<html>
<head><TITLE>Graphic Arts Publishing</TITLE></head>
<body><BODY BACKGROUND="parchmnt.gif">
<h2>Graphic Arts Publishing</h2>
<IMG SRC = "colorbar.GIF"><p>
<h3>Electronic Publishing Option</h3>

Combine theory with technological considerations and management skills. This option features
theoretical aspects of publishing and reproduction technologies, plus the practical side of soft-
ware and hardware applications and management strategies. Both corporate and commercial . .
```



Source Code

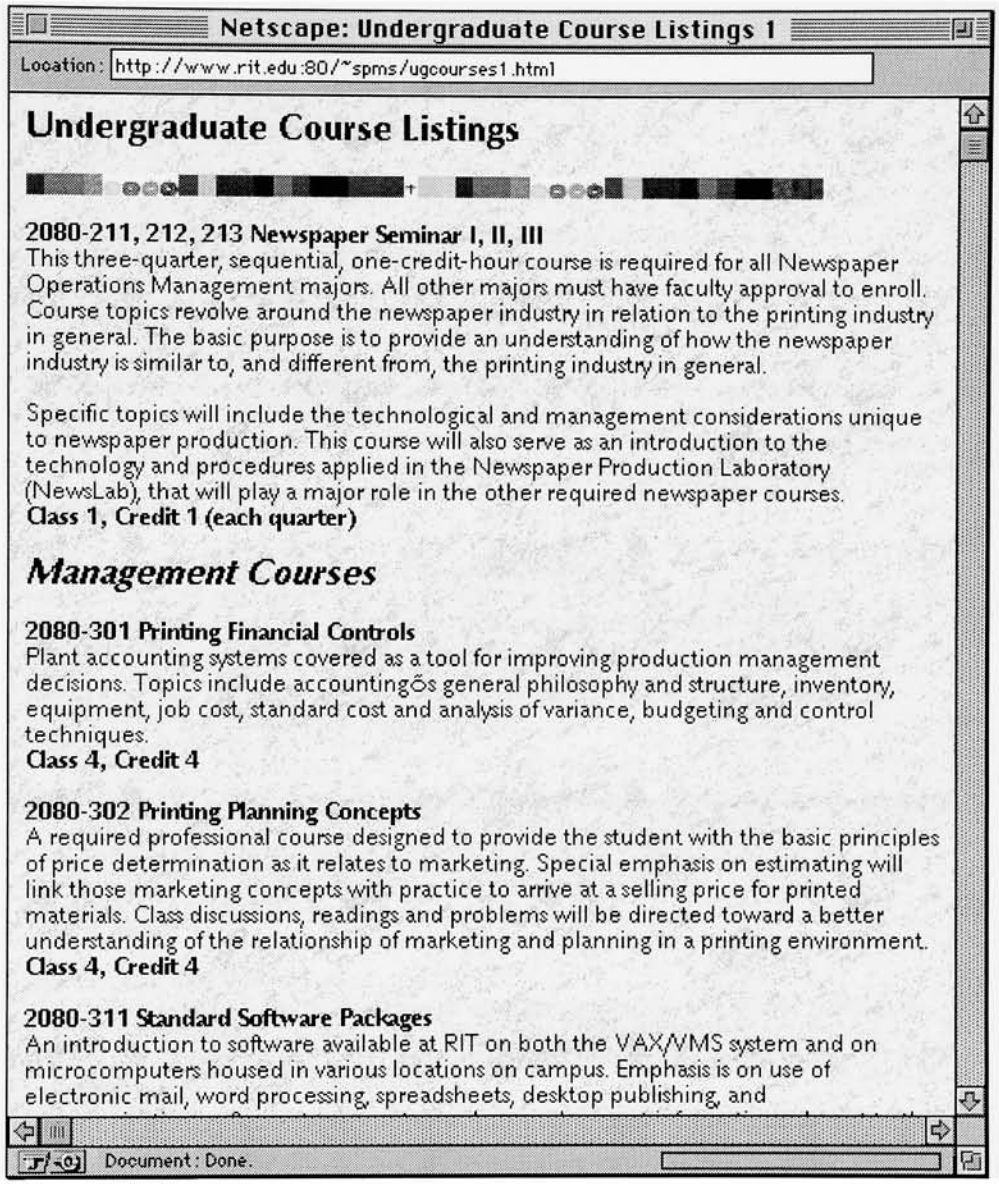
```
<html>
<head><TITLE>Graphic Arts Systems</TITLE></head>
<body><BODY BACKGROUND="parchmnt.gif">
<h2>Graphic Arts Systems</h2>
<IMG SRC = "colorbar.GIF"><p>
```

Acquire the spectrum of leadership skills needed to manage the technology of print communications. Topics such as financial and cost controls, marketing, and operations management are covered-plus the broad technological overview needed by an effective . . .



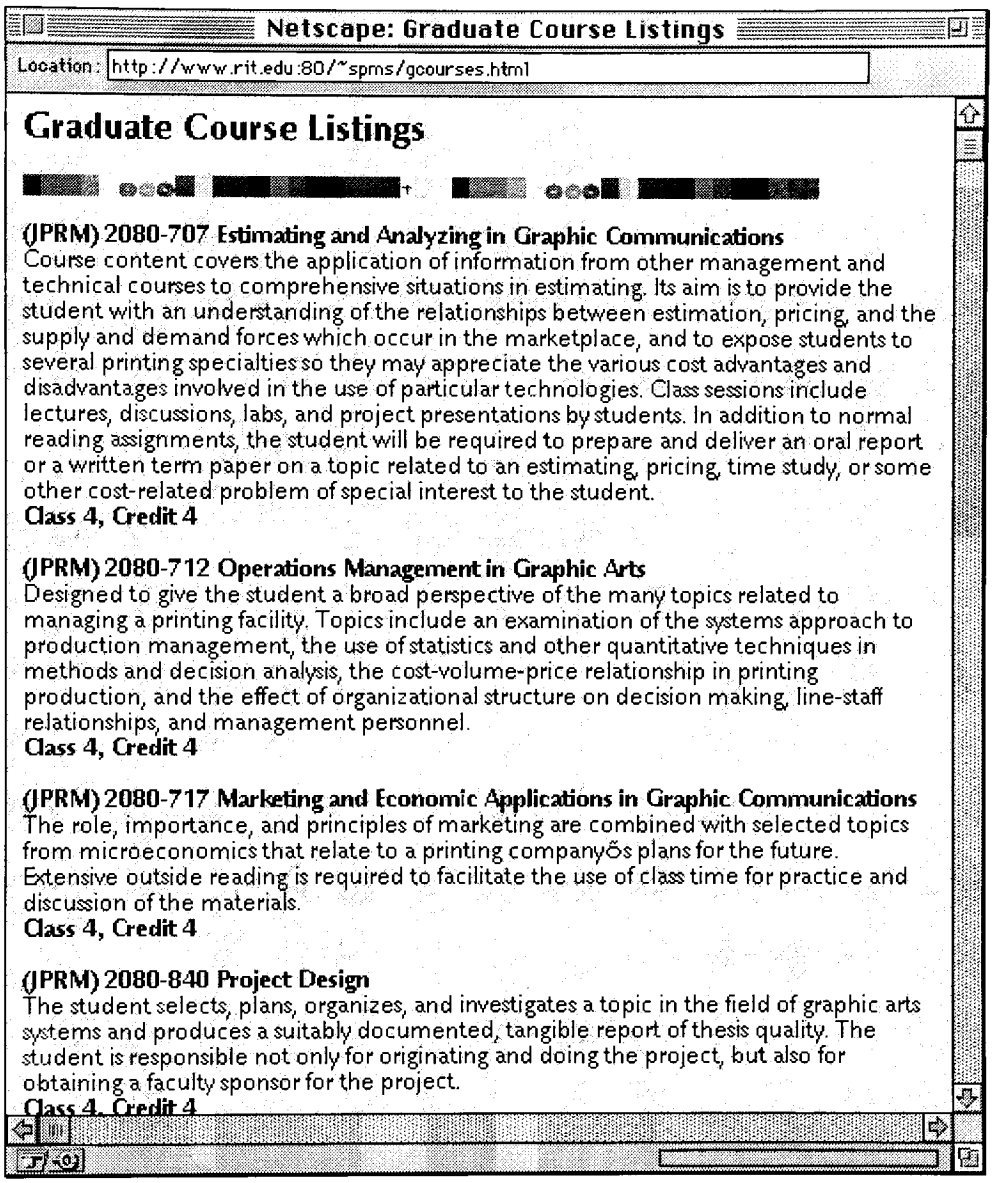
Source Code

```
<html>
<head><TITLE>Printing Technology</TITLE></head>
<body><BODY BACKGROUND="parchmnt.gif">

<h2>Printing Technology</h2>
<IMG SRC = "colorbar.GIF"><p>
Printing technology offers students an in-depth understanding of the technical concepts of imaging and an exploration of the technologies in demand by the computer industry. Graduates manage print production, do research, teach, and conduct technical training regardless of the medium. . .
```



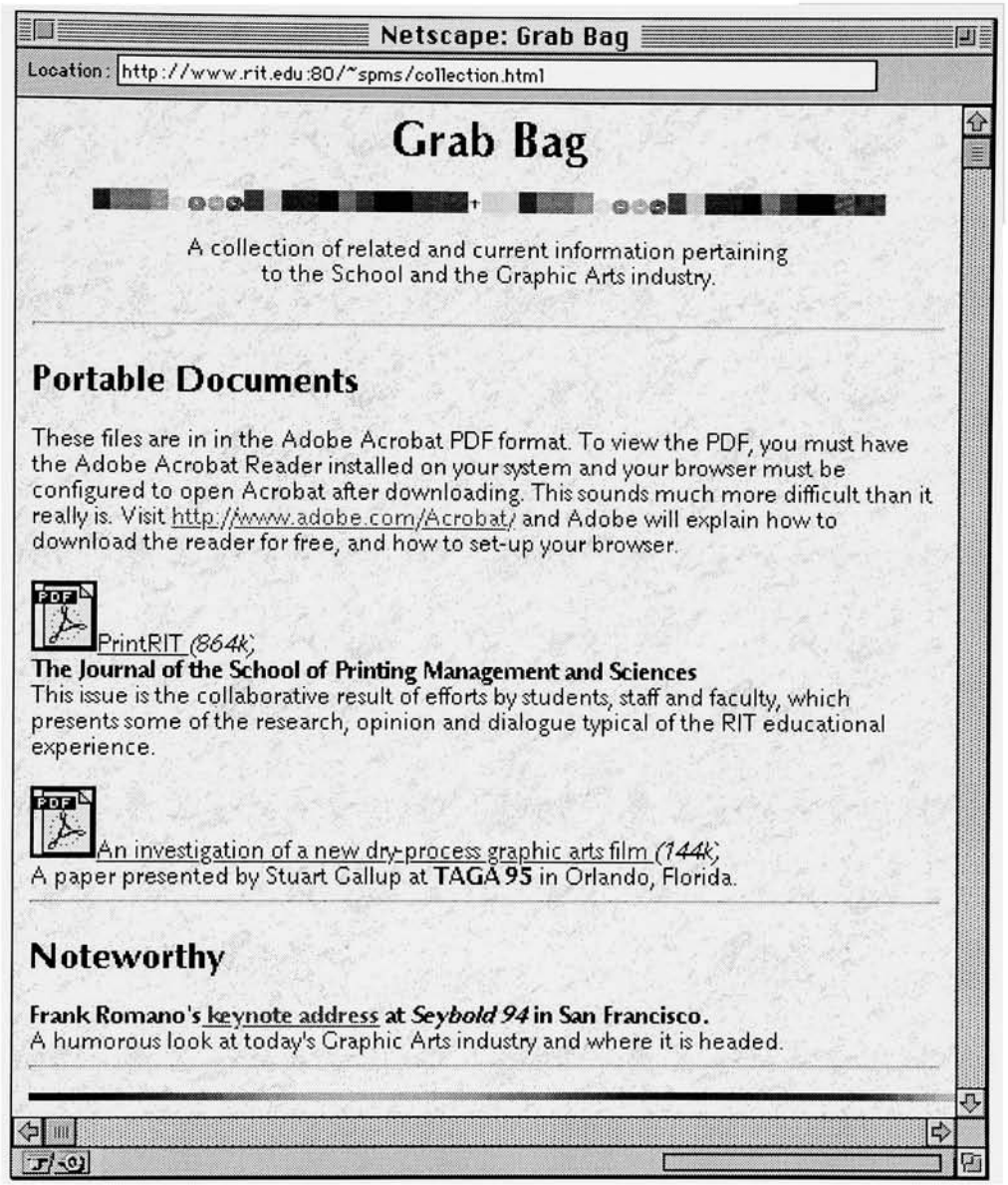
Source Code

```
<html>
<head><TITLE>Undergraduate Course Listings 1</TITLE></head>
<body><BODY BACKGROUND="parchmnt.gif">
<h2>Undergraduate Course Listings</h2>
<IMG SRC = "colorbar.GIF"><p>
<b>2080-211, 212, 213 Newspaper Seminar I, II, III</b><br>
This three-quarter, sequential, one-credit-hour course is required for all Newspaper Operations
Management majors. All other majors must have faculty approval to enroll. Course topics
revolve around the newspaper industry in relation to the printing industry in general. The basic
purpose is to provide an understanding of how the newspaper industry is similar to, . . .
```



Source Code

```
<html>
<head><TITLE>Graduate Course Listings</TITLE></head>
<body><BODY BACKGROUND="parchmnt.gif">
<h2>Graduate Course Listings</h2>
<IMG SRC = "colorbar.GIF"><p>

<b>(JPRM) 2080-707     Estimating and Analyzing in Graphic Communications</b><br>
Course content covers the application of information from other management and technical
courses to comprehensive situations in estimating. Its aim is to provide the student with an
understanding of the relationships between estimation, pricing, and the supply and demand . .
```



Source Code

```
<html>
<head><TITLE>Grab Bag</TITLE></head>
<body>
<BODY BACKGROUND="parchmnt.gif">
<center><h1>Grab Bag</h1>
<IMG SRC = "colorbar.GIF"><p>
A collection of related and current information pertaining <br>to the School and the Graphic
Arts industry.<p>
</center><p>
<hr> . . .
```

Appendix C

Appendix C

Eric J. Cohen
ROCHESTER INSTITUTE OF TECHNOLOGY

Electronic Publisher's Guide to the **World Wide Web**

Chapter 1
WHAT IS HTML

HTML 1
Tags 1
HTML and the appearance of a document 1
Structure 2
Any ASCII text editor 2

Chapter 2
BASIC STRUCTURE OF HTML

Basic structure of an HTML document 3
Exercise #1
Creating a basic HTML document 4
Tags covered in this chapter 4

Chapter 3
BASIC HTML TAGS

What are Headings? 5
Controlling text spacing 6
Exercise #2
Creating a real HTML document 7
Tags covered in this chapter 9

Chapter 4
CHARACTER STYLES

Character emphasis 10
Logical styles 10
Physical styles 12
Other formatting markup tags 13
Tags covered in this chapter 15

Chapter 5
LISTS

Unordered list 16
Numerically ordered list 16
Directory list 17
Definition lists 18
Tags covered in this chapter 18

Chapter 6

LINKS

Hot link 19
Hot link concepts 19
The link tag 20
Absolute vs Relative URLs 20
Excercise #3
Creating links 21
Tags/Attributes covered in this chapter 22

Chapter 7

GRAPHICS

Inline images 23
GIF 23
JPEG 24
Placing inline images in documents 24
Alignment 25
Image alternatives 27
Performance 28
External images 28
Images as links 29
Tags/Attributes covered in this chapter 30

Chapter 8

EXTERNAL MEDIA (VIDEO & SOUND)

Formats 31
Size and format indicators 32
Excercise #4
Images, video & sound 32
Data size 33

Glossary 34

WHAT IS HTML

HTML

HTML, or HyperText Markup Language is designed to specify the logical organization of a document, with important hypertext extensions. It is not a WYSWYG word processor such as Word or WordPerfect. This is because the same document may be viewed by many different browsers, of very different abilities. Thus, for example, HTML allows you to mark titles or paragraph marks, and then leaves the interpretation of these marked elements up to the browser. For example one browser may indent the beginning of a paragraph, while another may only leave a blank line.

HyperText Markup Language, or HTML requires no knowledge of programming to use, and is extremely simple to use in its basic form. A HTML document consists of nothing more than a plain text file with simple codes inserted into it. These codes, or "tags", instruct the viewing program (most likely a Web browser) to format sections of text, display images, and create hotlinks (connections) to other documents and files. No programming is involved at any point.

TAGS

A HTML document is formatted by enclosing sections of a plain text file between opening and closing tags. This is an example of a tag: <TAG>. Most tags come in pairs, with the closing tag preceded by a forward slash "/", like this: </TAG>. Tags define how the section of text they enclose will be displayed, create connections to other documents and files, such as multimedia documents like video clips and sound files, and cause images to be displayed within a document.

HTML AND THE APPEARANCE OF A DOCUMENT

The final appearance of a document is dependent on the program used to view it, the browser. Depending on the Web browser used, the final appearance of a document may vary significantly.

A document, for example, may be viewed using a text mode browser (such as Lynx), or a graphical browser (such as Netscape®). In the latter case, it may occupy the entire screen, or just a portion, be wider than it is tall, or vice versa, top level headings may be displayed using an extremely large font, or be only slightly larger than the text following, but heavily bold.

This is called "logical" markup: instead of specifying individual fonts, sizes and spacing, the designer of a HTML document inserts tags which indicate the "logical" representation of the document. For example, to create a heading, tags are inserted around it saying "this is a level one heading", instead of "this piece of text should be displayed in 18 point Helvetica bold with 24 point leading". The details are left up to the browser.

STRUCTURE

HTML identifies the structure of the document and it suggests the layout of the document. The display capabilities of the Web browser determine the appearance of the HTML document on the screen.

Using HTML you can identify:

- The title of the document.
- The hierarchical structure of the document with header levels and section names.
- Bulleted, numbered, and nested lists.
- Insertion points for graphics.
- Special emphasis for key words or phrases.
- Preformatted areas of the document.
- Hyperlinks and associated URLs.

HTML cannot control the:

- Typeface used for any document component.
- Point size of any specific font.
- Width or height of the screen.
- Centering, spacing, or line breaks of information, except in preformatted text.
- Background, foreground, or highlight colors.

These variables all depend on the browser.

ANY ASCII TEXT-EDITOR

To write an HTML document, all that is really needed is a basic ASCII test editor (for example, TeachText on the Macintosh), a full featured word processor or document layout software. The only requirement is that the software can save text in basic ASCII format. All HTML files must be saved with the appropriate .html extension (.htm for DOS)

BASIC STRUCTURE OF HTML

Basic Structure of an HTML Document

HTML documents are divided into two sections: a "head", and a "body". The "head" contains information about the document itself, such as the title, indexing information, and ownership. The "body" consists of the document itself, the images, text, and hotlinks to other documents and files displayed by the Web browser.

Head Section

```
<HTML>
<HEAD>
<TITLE>Sample Web Page</TITLE>
</HEAD>
```

The head section is opened and closed by the <HEAD> and </HEAD> tags. Information belonging in the head section should be placed between these two tags. The tags in the example above form a minimal head section, and should be included at the beginning of every HTML document.

Body Section

```
</HEAD>
<BODY>
<H1>Introduction</H1>
```

The body section is opened and closed by the <BODY> and </BODY> tags. The text and images of the document itself, to be displayed by the Web browser, should be placed within these two tags.

The opening tag should be placed immediately after the closing head tag (</HEAD>), and before any images or text included in the document itself. A matching </BODY> must be placed at the foot of the page. This tag ends the body section, and should be placed at the end of the document, after any images or text included in it.

Tags Ending a Page

```
</BODY>
</HTML>
```

</BODY> ends the portion of the page containing the document itself, and should be placed after any text or images that will be displayed

by the Web browser. </HTML> ends the page itself. The head and body sections are enclosed within the opening <HTML> and closing </HTML> tags.

EXERCISE #1
Creating a Basic HTML Document

Open your ASCII text editor of choice and type in the following code. Do not worry what <H1> means, you will learn what it is in the next chapter. Once you are done typing, save the document with the appropriate .html (.htm for DOS) extension; for example, `sample.html`.

Input

```
<HTML>
<HEAD>
<TITLE>Exercise 1</TITLE>
</HEAD>
<BODY>
<H1>Introduction</H1>
</BODY>
</HTML>
```

Open up your Web browser of choice and select the function which opens local files (**File/Open File** in Netscape®). The graphical representation displayed by the Web browser should look something like **Figure 1.** If it doesn't, look over the code and fix any mistakes. Resave the `exercise1.html` file and select **View/Reload** or its equivalent within the Web browser.

Output

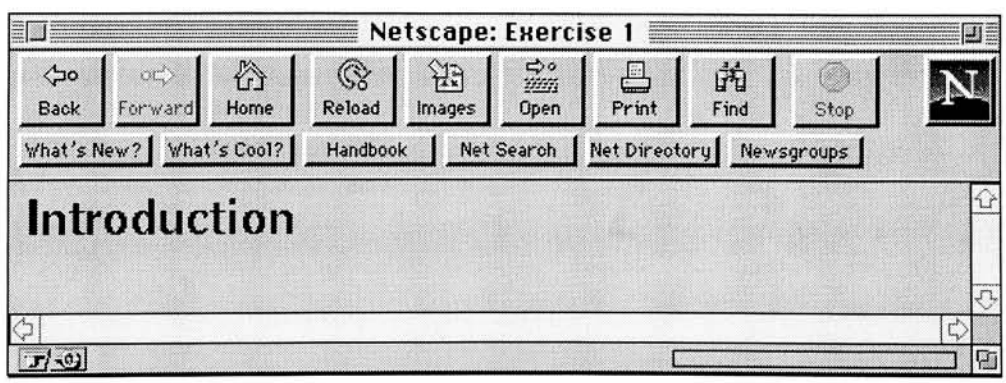


Figure 1: *The Output in Netscape®*

TAGS COVERED IN THIS CHAPTER

<HTML> . . . </HTML>	Document is HTML format
<HEAD> . . . </HEAD>	Head is enclosed in
<BODY> . . . </BODY>	Body is enclosed in
<TITLE> . . . </TITLE>	Title of document

WHAT ARE HEADINGS?

There are six levels of headings. The larger the number, the smaller the size of the resulting text (using a graphical browser). Examples of all six levels are provided below, followed by the tags that produce them. Headings are used to logically organize a page into separate sections for easier comprehension. Lower level headings in turn subdivide each section. Headings automatically create space between them and any text preceding or following them.

`<H1></H1>` . . . . . .Level One Heading
This is the highest level heading. It is often used at the beginning of a document.

`<H2></H2>` . . . . . .Level Two Heading
`<H3></H3>` . . . . . .Level Three Heading
`<H4></H4>` . . . . . .Level Four Heading
`<H5></H5>` . . . . . .Level Five Heading
`<H6></H6>` . . . . . .Level Six Heading

Input

```
<HTML>
<HEAD>
<TITLE>Headings</TITLE>
</HEAD>
<BODY>
<H1>Level 1 Heading</H1>
<H2>Level 2 Heading</H2>
<H3>Level 3 Heading</H3>
<H4>Level 4 Heading</H4>
<H5>Level 5 Heading</H5>
<H6>Level 6 Heading</H6>
</BODY>
</HTML>
```

Output

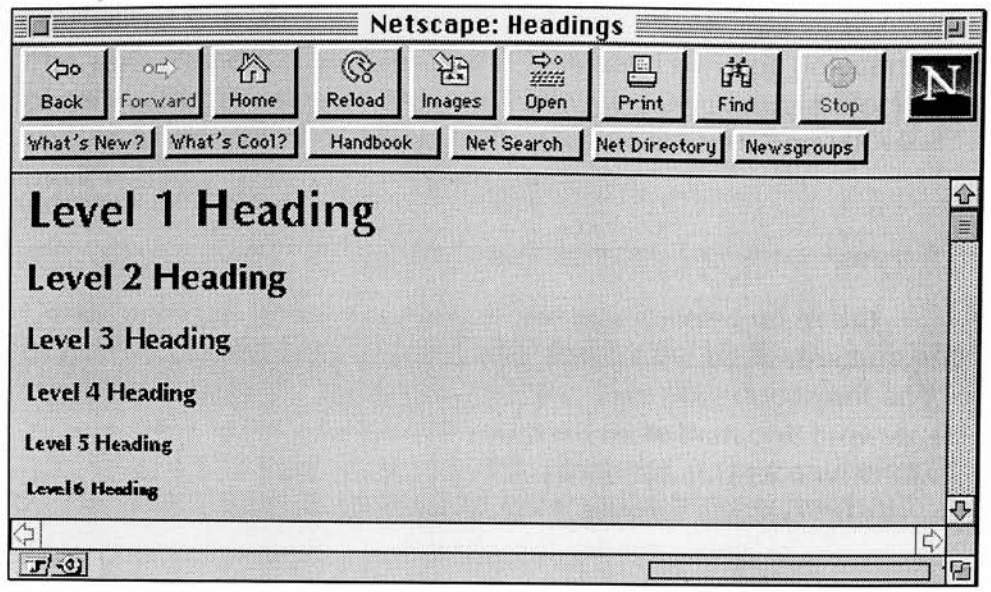


Figure 2: *The Output in Netscape®*

CONTROLLING TEXT SPACING

HTML leaves the details of how to display a document (spacing, font selection and size, etc.) up to the program viewing the document. White space (such as tabs, spaces, and carriage returns or line feeds) is ignored. Therefore, specific tags marking paragraph and line breaks must be inserted into HTML documents.

Paragraph Tags

The <P> tag starts a paragraph, and the </P> tag ends it. These tags create a space between any images or text preceding the enclosed text (some browsers also create a space after the closing paragraph tag).

```
<P>
This is an example.
</P>
```

Line Break Tag

The line break tag causes any text following it to begin on the following line.

```
This is<BR>
an example.
```

Horizontal Rule

The <HR> element is used to draw a horizontal dividing line completely across the screen. This can be to logically separate blocks of text, or to separate icon list from the body of the text. The HR element is empty (you don't need a </HR>).

EXERCISE #2
Creating a Real HTML Document

Create a simple HTML file using all the tags covered up to this point. You should have a comfortable feel for the basic structure and tagging schemes for creating HTML files. Save your document and view it under a Web browser. If you have more than one browser on your platform or have access to other platforms, it is recommended to view the document under each. Each browser displays HTML slightly differently and the changes may be surprising.

Input

```
<HTML>
<HEAD>
<TITLE>Exercise 2</TITLE>
</HEAD>
<BODY>
<H1>A Primitives Portfolio</H1>
<H3>By Malcolm Jones Jr.</H3>
<HR>
<P>
The one sure thing about Bob Dylan is that there is no
sure thing. In a musical career stretching over more than
three decades, he has proven time and again that he owns
the most bottomless bag of tricks in the business. . .
</P>
<HR>
From Newsweek<br>
March 20, 1995
</BODY>
</HTML>
```

Output

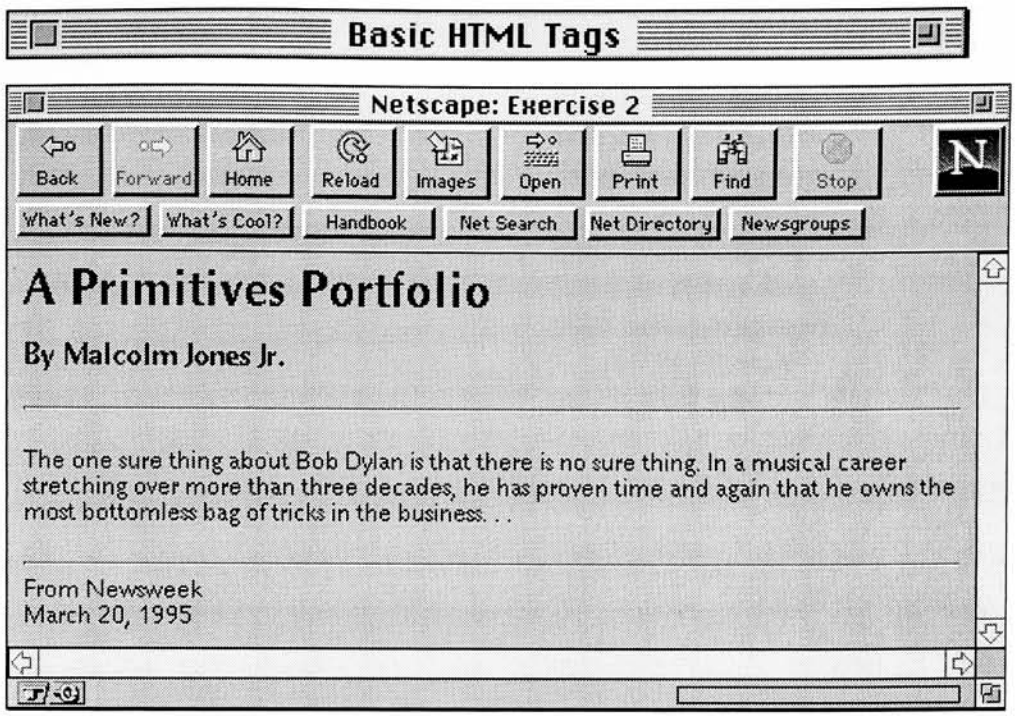


Figure 3: *The Output in Netscape®*

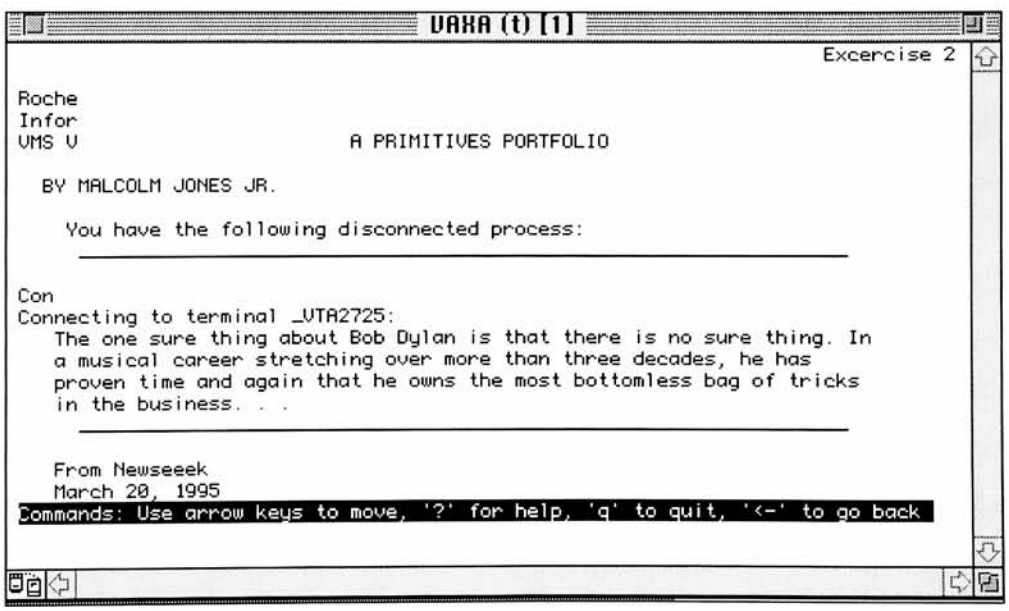


Figure 4: *The Output in Lynx*

Tags Covered in this Chapter

<H1> . . . </H1>	First level heading
<H2> . . . </H2>	Second level heading
<H3> . . . </H3>	Third level heading
<H4> . . . </H4>	Fourth level heading
<H5> . . . </H5>	Fifth level heading
<H6> . . . </H6>	Sixth level heading
<P> . . . </P>	Start new paragraph
  . . . </BR>	Forced line break
<HR>	Horizontal rule

CHARACTER STYLES

CHARACTER EMPHASIS

HTML allows you to specify special character highlighting or emphasis (boldface, italics, etc). These elements simply change the rendering of the characters enclosed inside the tags.

HTML allows you to specify these modes in two ways: Logically (by the logical meaning of the special text you wish to mark) and physically (by explicitly specifying the style you want, such as italics, boldfaced, underline, etc.). Note that the logical styles may not be distinct (i.e. different logical styles may be rendered in the same way).

LOGICAL STYLES

The logical styles and examples of their renderings:

EM—Emphasis (usually italics).

```
<EM> this is example text </EM>
```

is rendered:

"this is example text"

STRONG—Stronger emphasis (usually bold).

```
<STRONG> this is example strong text </STRONG>
```

is rendered:

"this is example strong text"

CODE—Example of typed code (usually fixed-width font).

```
<CODE> this is example code </CODE>
```

is rendered:

```
"this is example code"
```

SAMP—A sequence of literal characters.

```
<SAMP> this is example text </SAMP>
```

is rendered:

"this is example text"

VAR—A variable name.

```
<VAR> this is example text </VAR>
```

is rendered:

"this is example text"

DFN—The defining instance of a term (often rendered bold)

```
<DFN> this is example text </DFN>
```

is rendered:

"this is example text"

CITE—A citation (typically rendered in italics).

```
<CITE> this is example text</CITE>
```

is rendered:

"this is example text"

Physical Styles

The physical style elements, and their renderings, are:

TT—Fixed width typewriter font.

```
<TT> this is example text </TT>
```

is rendered:

"this is example text"

B—Boldface where available.

```
<B> this is example text </B>
```

is rendered:

"this is example text"

I—Italics

```
<I> this is example text </I>
```

is rendered:

"this is example text"

U—Underline (may be rendered as italic in some cases)

```
<U> this is example text </U>
```

is rendered:

<u>"this is example text"</u>

Other Formatting Markup Tags

If a browser does not support these tags, the text will appear normal. Browsers are programmed to ignore (and not display) tags they don't understand. Three of the most commonly supported examples are provided below.

ADDRESS

This element is used for address information, signatures, authorship, etc. The rendering of the contents is left up to the browser (indented, italic, right justified).

```
<ADDRESS> spms@rit.edu </ADDRESS>
```

is rendered:

spms@rit.edu

PREFORMATTED TEXT

The PRE element is used to enclose text to be displayed with a fixed width typewriter-like font. This is useful for presenting text that has been formatted for a fixed width character display.

Input

```
<PRE>
Chapter I       The Period
Chapter II      The Mail
Chapter III     The Night Shadows
Chapter IV      The Preparation
Chapter V       The Wine-shop
Chapter VI      The Shoemaker
</PRE>
```

Output

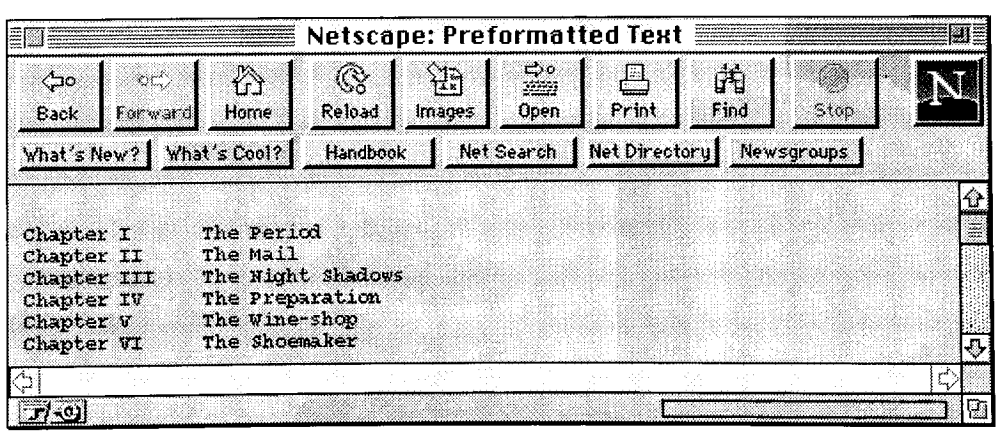


Figure 5: *The Output in Netscape®*

BLOCKQUOTE

The BLOCKQUOTE element allows quoted text to be rendered in an appropriate way. Typically this may be slightly indented, or italicized. BLOCKQUOTE also causes a paragraph break, and typically forces white space both before and after the quotation.

Input

```
<HTML><HEAD>
<TITLE>BLOCKQUOTE</TITLE>
</HEAD> <BODY>
<H1>A Tale of Two Cities</H1><HR>
<I>by Charles Dickens</I><BR>
A story of the French Revolution
<BLOCKQUOTE>
It was the best of times, it was the worst of times, it
was the age of wisdom, it was the age of foolishness, it
was the epoch of belief, it was the epoch of incredulity,
it was the season of Light, it was the season of
Darkness,it was the spring of hope, it was the winter of
despair, we had everything before us, we had nothing
before us, we were all going direct to Heaven, we were all
going direct the other way--in short, the period was so
far like the present period, that some of its noisiest
authorities insisted on its being received, for good or
for evil, in the superlative degree of comparison only.
</BLOCKQUOTE>
</BODY></HTML>
```

Output

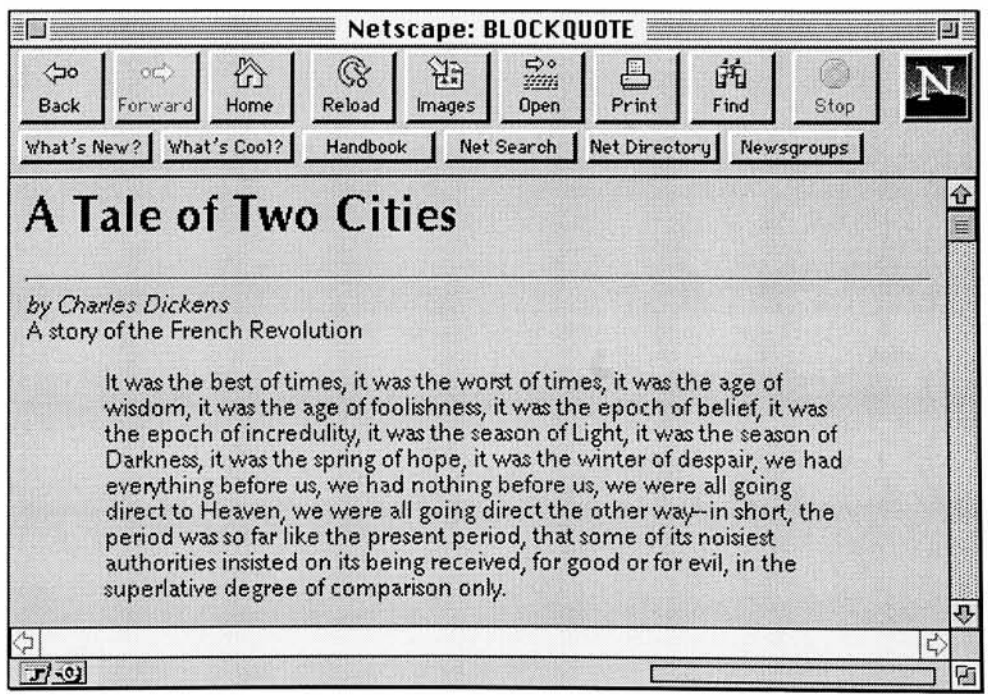


Figure 6: *The Output in Netscape®*

Tags Covered in this Chapter

`<EM> . . . </EM>`	Emphasized text
`<STRONG> . . . </STRONG>`	Strongly emphasized text
`<CODE> . . . </CODE>`	A code sample
`<VAR> . . . </VAR>`	A variable name
`<DFN> . . . </DFN>`	A term about to be defined
`<CITE> . . . </CITE>`	A citation
`<TT> . . . </T>`	Typewriter text
`<B> . . . </B>`	Bold text
`<I> . . . </I>`	Italicized text
`<BLOCKQUOTE> . . . </BLOCKQUOTE>`	A long quotation
`<ADDRESS> . . . </ADDRESS>`	A signature
`<PRE> . . . </PRE>`	Preformatted text

List tags cause standard items (such as bullets or numbers) to be inserted before the listed text. They can be "nested" within one another to create sub-lists (the effect of this on the appearance of a document differs with the browser used). The text of a list entry uses the same style and fonts as normal text.

UNORDERED LIST

Using a unordered list tag causes bullets (or the equivalent, if the viewer is using a text mode browser such as Lynx) to be placed before the listed text. An "unordered list" is opened by <UL> and closed by </UL>. Listed items between them are preceded by <LI>.

Input

```
<UL>
<LH>This is a list header!
<LI>Any
<LI>Text
<LI>Here
</UL>
```

Output

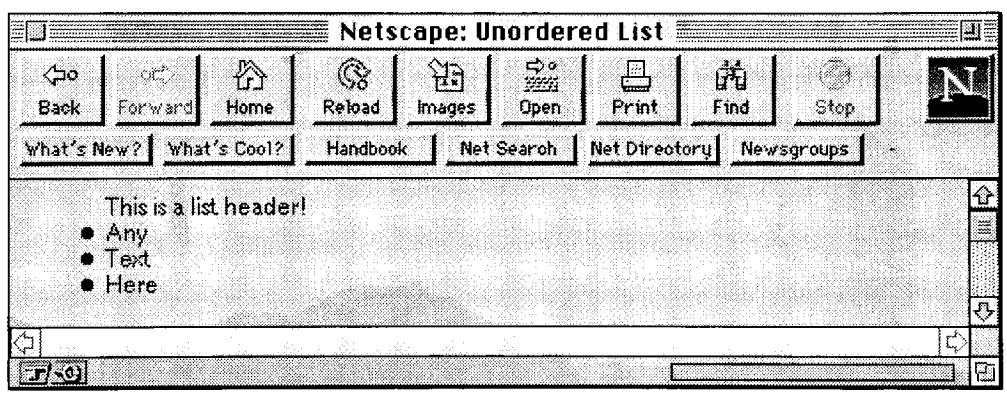


Figure 7: *The Output in Netscape®*

NUMERICALLY ORDERED LIST

Using a numerically ordered list tag causes numbers (starting from one) to be placed before the listed text. An ordered list is opened by <OL> and closed by </OL>. Listed items between them are preceded by <LI>.

Input

```
<OL>
<LH>This is a list header!
<LI>One
<LI>Two
<LI>Three
</OL>
```

Output

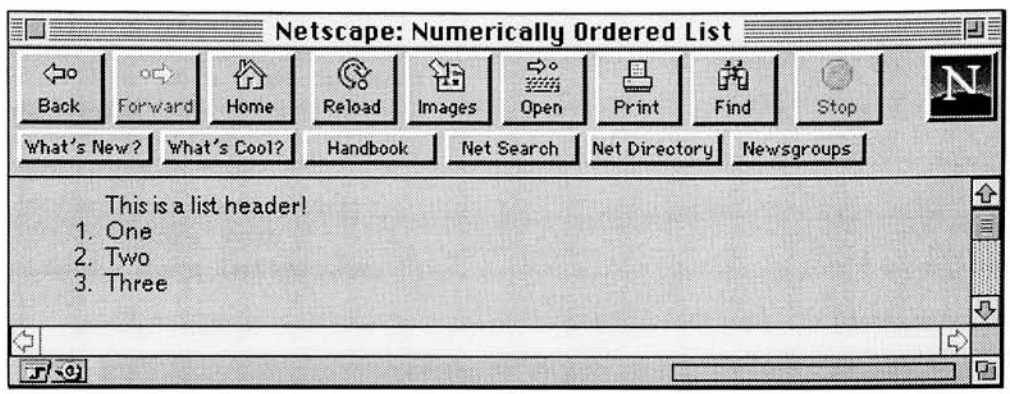


Figure 8: *The Output in Netscape®*

DIRECTORY LIST

This is used for short lists of items, such as file names.

Input

```
<DIR>
<LI>alpha.html
<LI>beta.html
<LI>theta.html
<LI>zeta.html
</DIR>
```

Output

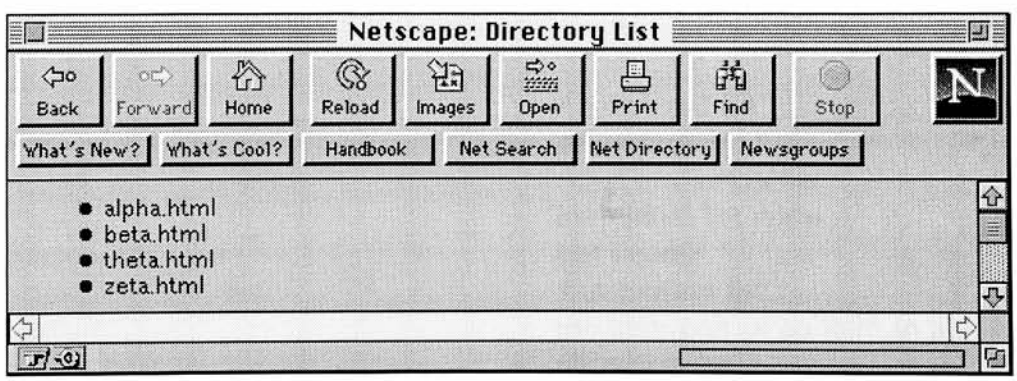


Figure 9: *The Output in Netscape®*

Definition List

Definition lists place nothing before entries preceded by <DT>, and indent entries following <DD>. An definition list is opened by <DL> and closed by </DL>. Listed items between them are preceded by <DT> (Definition Term) and then <DD> (Definition Definition). Use of <DT> without <DD> or vice versa is non-standard.

Input

```
<DL>
<LH>This is a list header!
<DT>Term
<DD>Definition
<DT>Another Term
<DD>Another Definition
<DT>One More Term
<DD>One More Definition
</DL>
```

Output

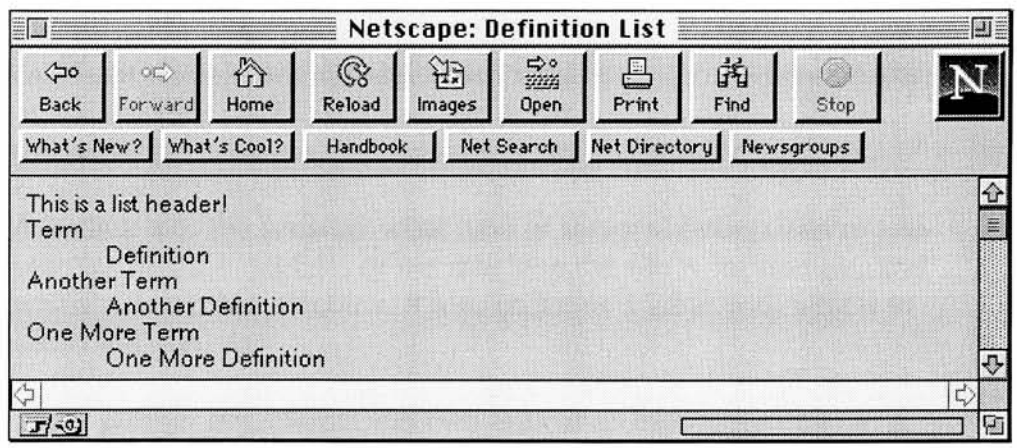


Figure 10: *The Output in Netscape®*

Tags Covered in this Chapter

<OL> . . . </OL>	An ordered list
<UL> . . . </UL>	An unordered list
<DIR> . . . </DIR>	A list of short items
<LI>	Individual list items
<LH>	List header
<DL> . . . </DL>	A definition list
<DT>	the term part of an item in a definition list
<DD>	The definition part of an item in a definition list

HOT LINK A hot link is a word, phrase or image which is highlighted by your browser indicating to the reader that more information on the highlighted item is available by clicking on that item. When you select or click on a hotlink, you are automatically transferred to the document (or image, etc.) which that hotlink points to (or if the selected resource is not something your browser knows inherently how to display, like a video for instance, an external viewer is launched which knows how to display the selected resource).

Hotlinks are created by placing a word, phrase, or image within a pair of anchors which associate the text with a **URL** (Uniform Resource Locator). A URL is simply the Internet address of another file. Image hotlinks are called clickable images. A complex image can even contain multiple hotlinks (depending upon where on the image the user clicks), using image maps.

HOT LINK CONCEPTS A hotlink tells the browser what type of server is being used to present the file, the address of the machine the file is located on, and where it is on that machine. It also includes a label with which to select the hotlink.

Links to other documents and files appear as highlighted text in a unique style when a page is viewed. When the person using the browser selects the link, the browser is automatically transferred to the file the link points to. The file can be located anywhere in the world.

This is the power of the World Wide Web: transparent linking of information resources to one another across the planet. The person viewing a document does not need to know how or where a resource is stored or what server is used to present it, they just select a hotlink and the file is automatically transferred to their machine. Three things can happen at this point:

1) The browser itself will display the file (if it is a text file, another HTML document, or a type of image the browser is capable of displaying).

2) If the browser is configured properly, and recognizes what type of file is being sent to it, a program capable of viewing the file will be launched and the incoming file will be automatically loaded into it (for example, a video or audio clip).

3) If the browser does not recognize what type of file is being sent to it, or it has not been told what type of program to use to view the incoming file, the user will be told this, and asked whether or not they want to download the file.

The Link Tag

When creating links in HTML documents the link tag `<A . . . /A>` is used. The most common kind of linking among HTML documents is linking to other documents on the Web. Unlike most other tags the `<A>` tag includes some extra information about the link itself, called attributes. To link to other documents on the Web, the `HREF` attribute of the Link tag is used. `HREF` is short for hypertext reference and is used to specify the URL of the file the link points to. Below is what a typical `<A>` tag looks like:

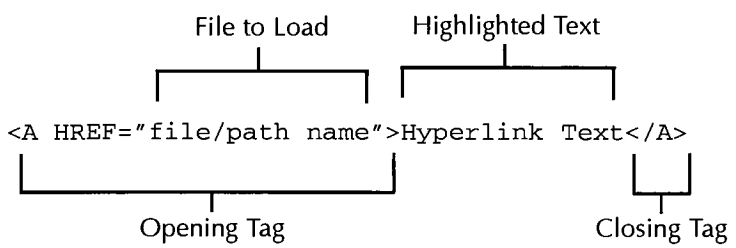


Figure 11: *Link Tag*

Absolute vs Relative URLs

A relative URL assumes the same access-method, server-name, and directory-path as the document the URL appears in. It indicates the relative position of the target URL from the current URL. For example, here's a URL to a file in the same directory as the document the URL appears in.

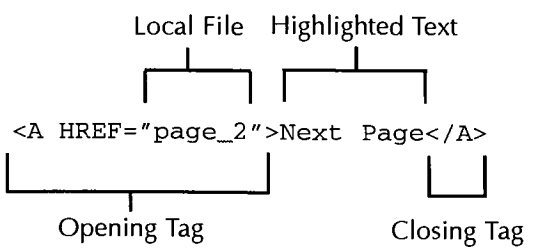


Figure 12: *Relative URL*

Relative URLs are also sometimes called partial URLs. They are used to point to information resources in the same directory or on the same server. Absolute URLs, on the other hand, are usually used to point to information resources on other servers. In practical terms, this means that you can use relative URLs to direct navigation between documents that you author and use absolute URLs to direct navigation to resources elsewhere on the Internet. A typical link with an absolute path looks like the one below:

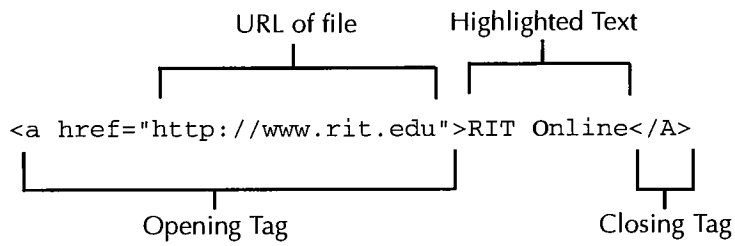


Figure 13: *Absolute URL*

EXERCISE #3
Creating Links

Exercise 3 demonstrates how to include both an absolute and relative link within an HTML document. For the relative path, try linking to Exercise 2. Remember that both exercises need to be in the same directory (folder) for the relative path to work properly. For the absolute path, link to any Web site you wish.

Input

```
<HTML><HEAD>
<TITLE>Linking</TITLE>
</HEAD>
<BODY>
<H1>Linking</H1>
<HR>
<H3>Relative and Absolute paths</H3>
This Relative path connects me to <A HREF="Exercise
2.html">Exercise 2</A>
<BR>
<BR>
This Absolute path connects me to <A
HREF="http://www.rit.edu">RIT Online</A>
</BODY>
</HTML>
```

Output

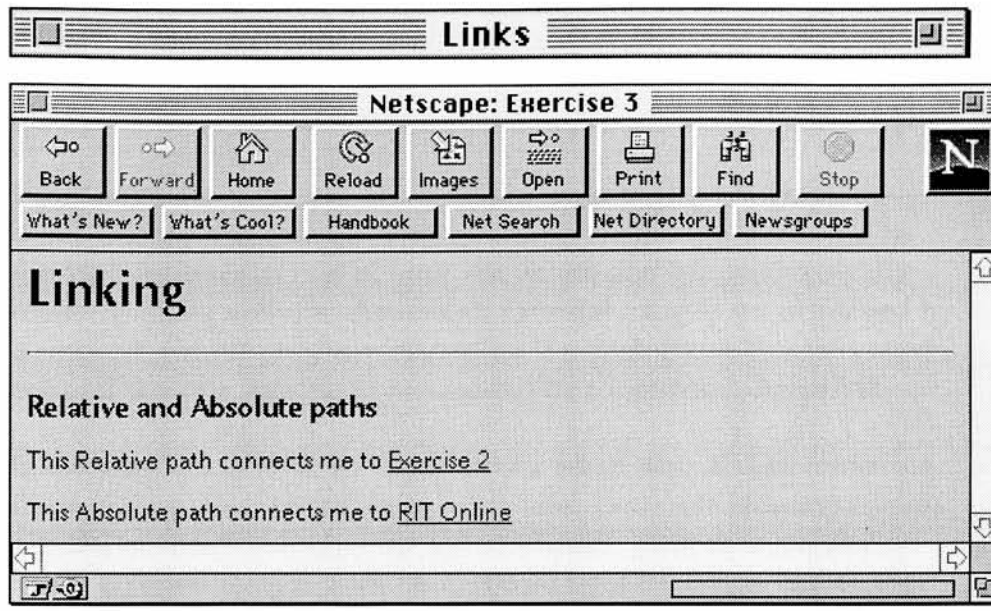


Figure 14: *The Output in Netscape®*

The distinction between absolute and relative URLs is totally hidden from the Web server. When a user selects a relative hyperlink, the Web browser uses the current URL to determine the access-method, server-name, and directory-path, and sends only absolute URLs to the server.

Relative URLs also give you the flexibility to move your HTML directory structure anywhere on the Web. This means if you move your documents to another server, all the linking remains intact.

TAGS/ATTRIBUTES COVERED IN THIS CHAPTER

`<A> . . . </A>`	Link tag
`<HREF>`	Hypertext reference

7

GRAPHICS

Fundamentally, there are two different ways to present graphics—inline images and external images. Inline images are displayed by the Web browser as part of your document and are automatically retrieved along with the document. External images are displayed by a separate viewer program (started by the Web browser when needed) and must be specifically requested by triggering a hyperlink.

Inline graphics involve the transfer of a lot of data and so retrieving them can be slow. Fortunately, many Web browsers let users optionally delay downloading of the inline images. With delayed downloading, inline images must be triggered by a hyperlink before they will be downloaded. When triggered, they still appear inline, not in an external viewer.

When contemplating the use of graphics within your HTML documents, you should consider the user community that will be accessing your Web server. Will they be using graphics-capable Web browsers, such as Netscape®, or will they be using line-oriented Web browsers, such as Lynx? The decision is easier if one class of Web browser dominates, but is more complicated if you expect people with both types of browsers to be using your Web server.

INLINE IMAGES

The inline image is just one more piece of information that is included in the autoflow and autowrap of your HTML on the Web browser screen. In other words, an image is treated just like a word. So an image can appear in the middle of a paragraph. If you want the inline image to stand alone you need to make sure you place <P> or
 HTML tags around it in your HTML document.

GIF

Although many different graphics file formats exist, most browsers will only recognize a few. In practice the only format common to all browsers is GIF (Graphic Interchange Format) devised by CompuServe. While this is the most commonly used format on the Web, other graphics file formats have their strengths and some look set to increase in popularity. Browsers that can display images in other formats, primarily JPEG (Joint Photographers Expert Group) are starting to appear. As yet such browsers are not in common use, but are likely to be so within the next year.

GIF was developed to be a device-independent method of storing pictures. GIF allows high-quality, high resolution graphics to be displayed on a variety of graphics hardware and is intended as an exchange and display mechanism for graphic images. GIF is reasonably well matched to inexpensive computer displays, since it can only store 8 bits/pixel (256 or fewer colors) and most PCs can not display more than 256 distinct colors at once. GIF does well on images with only a few distinct colors, such as line drawings and simple cartoons. A GIF picture file has an extension .GIF.

JPEG

The JPEG standard is an excellent standard for most realistic images (photos for example, but not line drawings or logos). It uses a powerful, though nominally "lossy", compression method. JPEG is best suited for truecolor original images.

JPEG stores full color information: 24 bits/pixel (16 million colors). Therefore, with full-color hardware, JPEG images look much better than GIFs on such hardware. JPEG files are much smaller than GIFs, therefore, they are superior to GIF in terms of disk space saving and transmission time. A JPEG picture file has an extension .JPG.

Using JPEG for a photographic image for example can produce 10:1 savings compared to GIF, as well as permitting much better display quality on truecolor-capable displays. Netscape® handles inline JPEG; most older browsers need to use an external JPEG viewer.

Placing Inline Images in Documents

Inline images are specified in HTML using the <IMG> tag. Like the <P> and
 tags, the <IMG> tag has no closing tag. It does however, have flexibility in that it can include up to three attributes; SRC, ALIGN and ALT.

The SRC attribute indicates the filename of the image to be included within the document. It is followed by an equals sign, with the filename in quotes. Be advised that filenames are case sensitive, so be sure to name the file exactly. Also, the image file should be in the same directory as the HTML file. A typical <IMG> tag looks like this:

```
<IMG SRC="filename.GIF">
```

Input

```
<HTML><HEAD>
<TITLE>Images</TITLE>
</HEAD>
<BODY>
<IMG SRC="WBradley.GIF">
</BODY>
</HTML>
```

Output

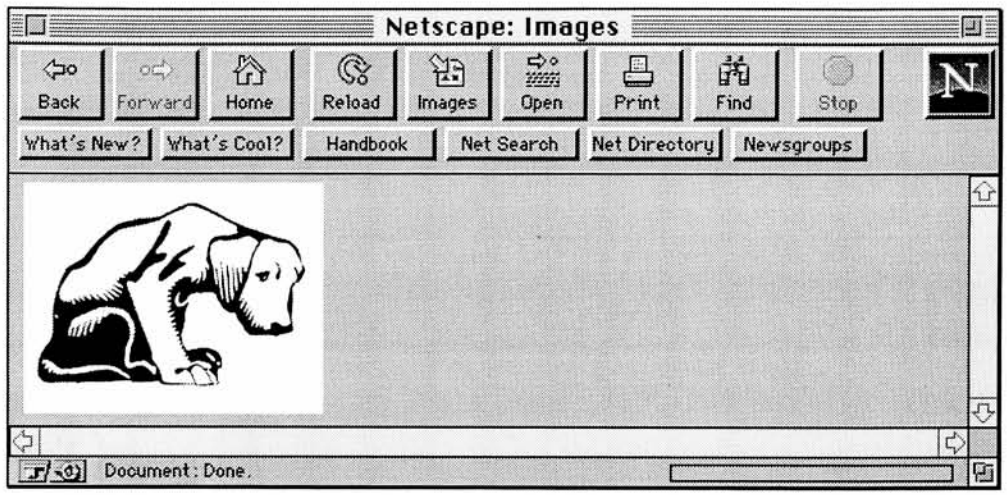


Figure 15: *The Output in Netscape®*

ALIGNMENT When inline images are autoflowing as part of a paragraph, you can explicitly control the alignment of the image with the text line by using the optional ALIGN attribute of the <IMG> tag. The three values for ALIGN are:

```
<IMG ALIGN=TOP SRC="filename.GIF">

<IMG ALIGN=MIDDLE SRC="filename.GIF">

<IMG ALIGN=BOTTOM SRC="filename.GIF">
```

TOP alignment places the top of the image even with the top of the current line of text, and so on. If ALIGN is omitted, bottom alignment is the default.

Input

```
<HTML><HEAD>
<TITLE>Image/Text Alignment</TITLE>
</HEAD>
<BODY>
<IMG SRC="Horse.GIF" ALIGN=TOP>Text at top of an image
<HR>
<IMG SRC="hglass.GIF" ALIGN=MIDDLE>Text in the middle of
an image
<HR>
<IMG SRC="doggy.GIF" ALIGN=BOTTOM>Text at bottom of an
image
</BODY>
</HTML>
```

Output

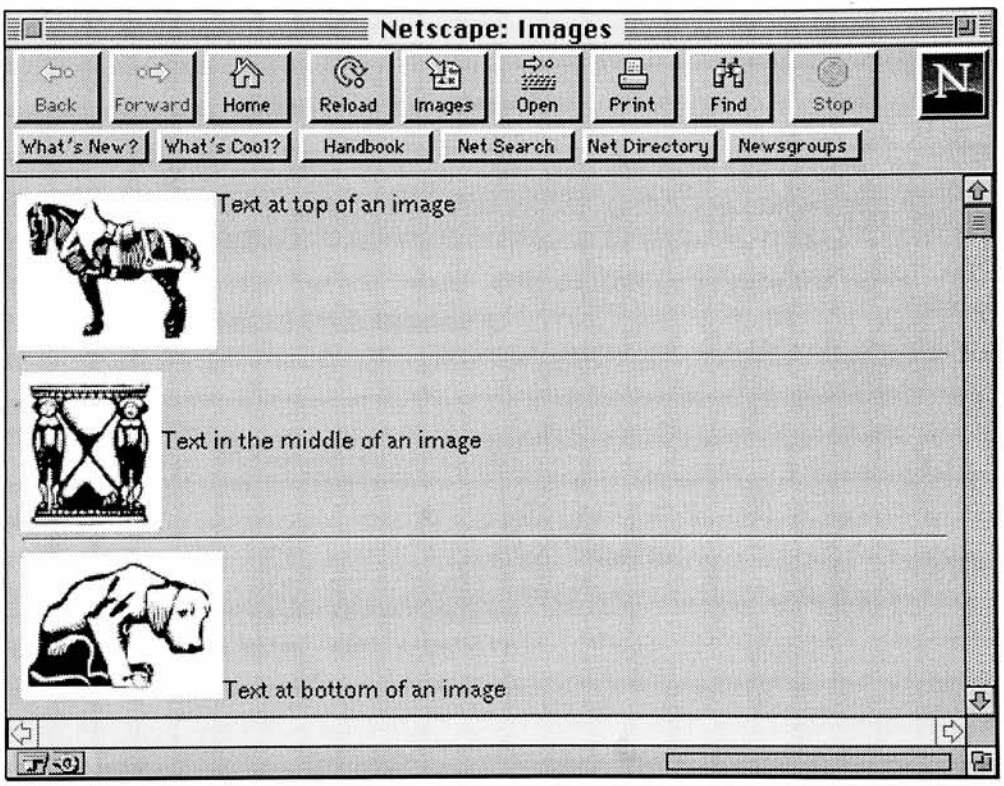


Figure 16: *The Output in Netscape®*

Image Alternatives

Inline images can't be displayed on character-based terminals. Character-based Web browsers indicate that an image is in the autoflow stream by displaying:

```
[IMAGE]
```

However, you can override this default and make your inline images more meaningful in character-based Web browsers by using the ALT attribute:

```
<IMG SRC="filename.GIF" ALT="image description">
```

The ALT option makes character-based Web browsers display the image description instead of [IMAGE]. You can also make the character-based browser ignore an image using a null ALT option:

```
<IMG SRC="filename.GIF" ALT="">
```

When a character-based Web browser sees the null ALT option, it ignores the image-insertion tag. The HTML authoring convention is that if you are going to insert inline images in your document, you should at least describe the image with a word or phrase for users with character-based Web browsers.

Input

```
<HTML><HEAD>
<TITLE>Image Alternative</TITLE>
</HEAD>
<BODY>
<IMG SRC="prev.GIF" ALT="[Previous Page]">
<IMG SRC="next.GIF" ALT="[Next Page]>
<IMG SRC="help.GIF" ALT="[Help]>
<IMG SRC="quit.GIF" ALT="[Quit]>
</BODY></HTML>
```

Output

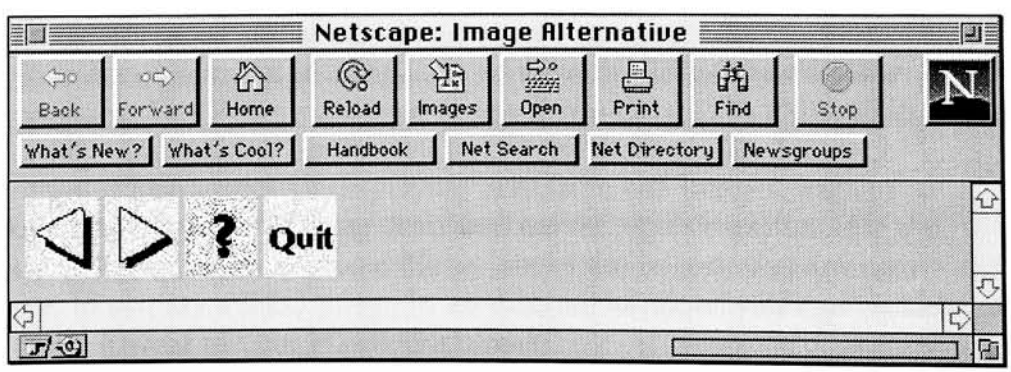


Figure 17: *The Output in Netscape®*

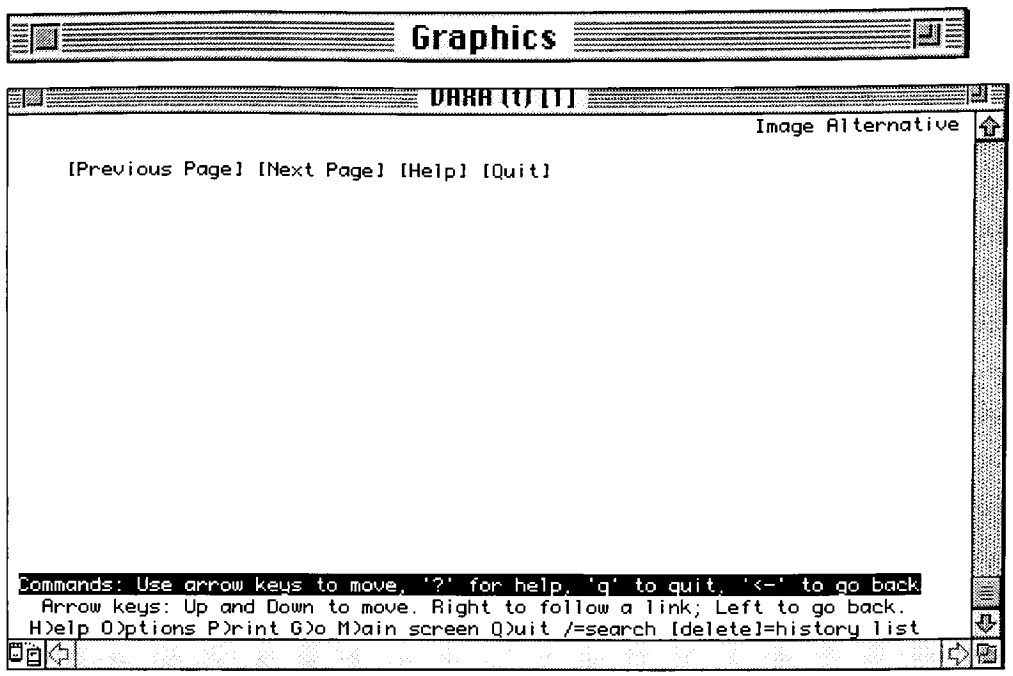


Figure 18: *The Output in Lynx*

PERFORMANCE Think about performance. It takes a lot longer for a Web browser to retrieve an HTML document that has inline images than to retrieve one that does not (unless the user specifies delayed downloading). The larger the inline image, the longer it takes. In fact, the time is proportional to the square of the dimension (a four-inch-square image takes almost twice as long as a three-inch-square one). One thing working in your favor is that many graphically-based Web browsers cache inline images. So if the inline image has been used in previous HTML documents in the user's navigation chain, that image might still be cached. Caching also happens when the same inline image appears multiple times in the same HTML document. It is retrieved only once.

EXTERNAL IMAGES External images are images that are not displayed inline as part of your document, but in a separate window, by an external viewer program. This is the technique to use if you have TIFF, JPEG, RGB, or HDF images and don't want to convert them to GIF. This technique is also useful for displaying very large GIF images. Instead of using the <IMG> HTML tag described earlier, you simply include the URL of your external image file as part of the hyperlink. For example, to display a JPEG image in an external window, you can include the following in your HTML document:

```
<A HREF="filename.GIF">hyperlink text</A>
```

The user needs an external viewer installed that knows how to deal with the incoming external image file. Also the Web browser needs to know how to recognize the type of data that's arriving and how to start the appropriate viewer.

Using external images with your HTML document further limits the usability of your HTML document. Just as there are users who can't display inline graphics, there are more who have not configured their environment to support the display of external images.

IMAGES AS LINKS

If you include an <IMG> tag inside of a link tag (<A>), the image specified becomes a clickable link. One trick that's used to solve the performance problem of big images is to use "thumbnails." A thumbnail is a small version of a figure displayed inline, which is a link to the full-sized image displayed externally.

Input

```
<HTML><HEAD>
<TITLE>Images as Links</TITLE>
</HEAD>
<BODY>
<A HREF="hi_res.GIF"><IMG SRC="thumb.GIF"></A>Click on
image to see high-res representation!
</BODY>
</HTML>
```

Output

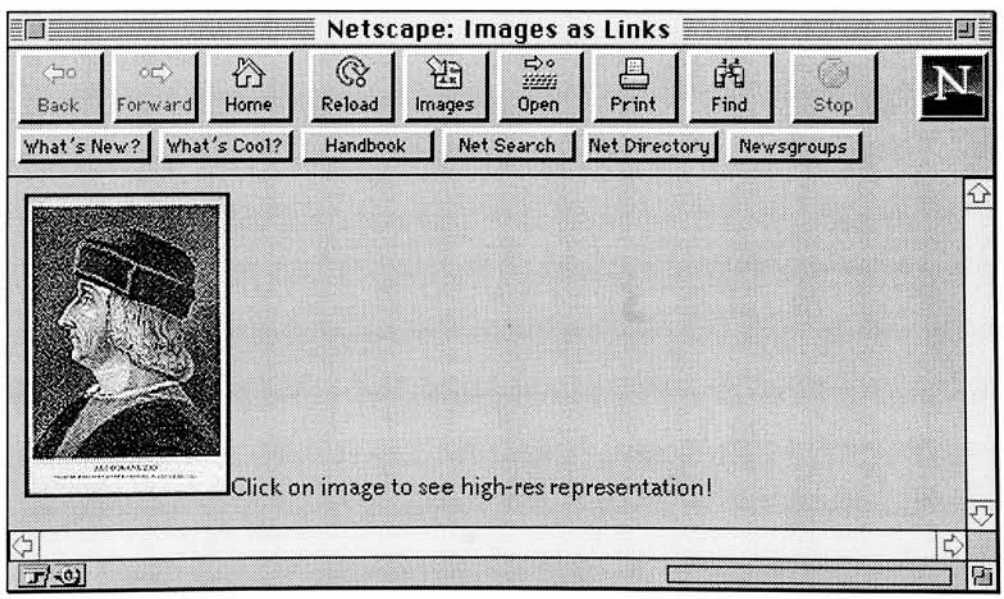


Figure 19: *The Output in Netscape®*

TAGS/ATTRIBUTES COVERED IN THIS CHAPTER		
	<IMG>	Image
	SRC	Filename of image
	ALIGN	Alignment of text next to image
	ALT	Character-based display

EXTERNAL MEDIA (Video & Sound)

To specify external media in HTML, you create a link to it just as you would another HTML document. In other words, you use an <A> tag with the HREF attribute.

```
<A HREF="external_file.ext">media file</A>
```

The file format of the external media is a **critical** issue when implementing them within an HTML document. The file format signals the the Web browser what type of file is about to be downloaded. The Web browser, if configured properly, then automatically launches an external program that plays the media file.

FORMATS

The best format for providing sound on the Web is the AU format. The reasons for this choice are: this format can be supported by most of the sound capable machines on the Internet. Secondly, this format produces a fairly small file that requires little bandwidth to transfer. AU files are of only barely acceptable quality, as the 8-bit sampling causes them to sound comparable to transmission over a telephone. Audio files in the AU format have an .AU extension.

Like other files, video files can be identified by their file extensions. There are only few movie file formats that can be viewed from the Internet, which are international standard file formats for multimedia. They are MPEG (Moving Pictures Expert Group) and Quicktime. MPEG movie formats have the extension .MPG. QuickTime movies have file extension .QT and .MOV. .MOV files can be played on both Macintosh and PCs, while .QT extensions can only be viewed on a Macintosh.

	Format	*Extension*
Audio	AU	.AU
Video	MPEG	.MPG
	QuickTime	.QT or .MOV

Figure 20: *File Formats and Extensions*

Size and Format Indicators

It is considered good etiquette to label all large files (larger than about 100K). This applies to graphics, audio, video, and other large file formats. Because some clients can only accept certain data formats, it also is helpful to provide descriptors or icons that identify a file as image, audio, or video and to indicate its data format and size. This is especially true of very large files. It is extremely frustrating for users to download a 5MB video file only to discover that it is in a format not supported by their browsers.

Exercise #4
Images, Video & Sound

Create an HTML document with the following characteristics:

- Utilize images as links.
- Include a video file.
- Include an audio file.

Input

```
<HTML><HEAD>
<TITLE>Exercise 4</TITLE>
</HEAD> <BODY>
<H1>Exercise 4</H1>
<HR>
<A HREF="sound.AU"><IMG SRC = "sound.GIF"  ALIGN=MIDDLE
ALT=" ">Sample Sound </A>[64k]  <I><B>AU Format</B></i>
<HR>
<A HREF="movie.mov"><IMG  SRC="movie.GIF" ALIGN=MIDDLE
ALT=" ">Sample Movie</A> [4.3MB]   <I><B>QuickTime
Format</B></i>
</BODY>
</HTML>
```

Output

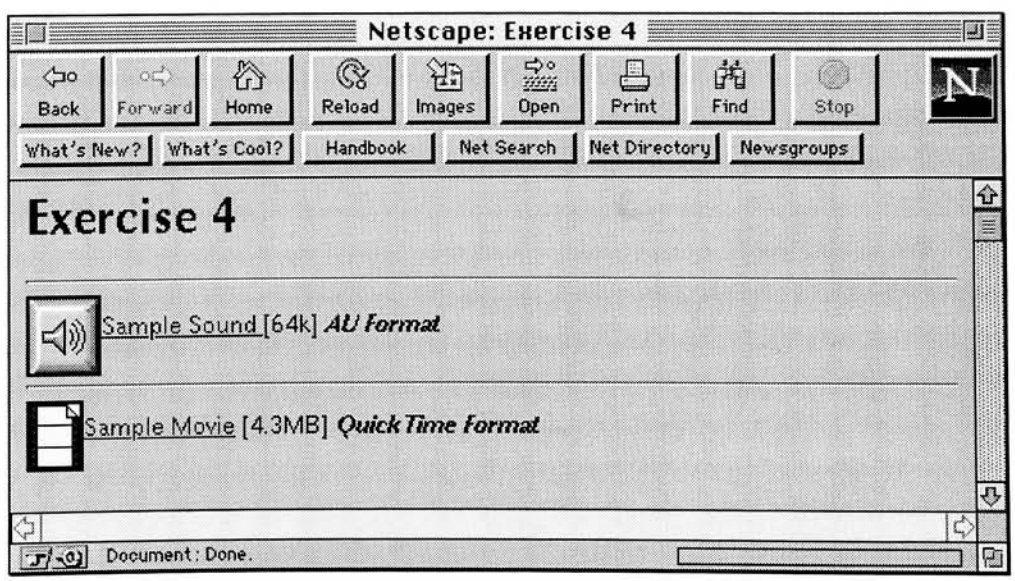


Figure 21: *The Output in Netscape®*

DATA SIZE The problem of data size can become significantly worse when dealing with audio and video files. A 10-second low-quality audio clip can take more than a minute to download over a 14.4-Kbps line. The size of higher-quality audio clips—in stereo or at higher-sampling rates—is at least double that of low-quality clips.

Depending on their quality and duration, video clips generally range from hundreds of kilobytes to 50MB or more. Even a few seconds of low-quality video—without audio—can cause significant waiting on a 56-Kbps line. For video as small as 30 seconds, may take minutes even over a full T1 (1.54 Mb/sec.) connection. For this reason, audio and video files should be used sparingly and should be marked carefully and appropriately as large files.

The best means of reducing the data size for audio or video files is to reduce their quality. For example, changing from stereo to mono sound can yield a data-size reduction as large as 50 percent. Lowering the frame rate or resolution in a video file or the sampling frequency in an audio file can substantially reduce the file data size.

Similarly, it is important to use reasonable-length video and audio clips (less than one minute). On some systems, it is difficult to stop a video or audio file from playing once it is started. Therefore, in addition to keeping the data size of the file small, use shorter clips to give your users greater control over what they have to view.

Browser:
Software that provides an interface to the World-Wide Web, such as Netscape or Mosaic

Graphics Interchange Format (GIF):
A commonly used graphics format which compresses the image and stores color information within the file.

Helper Application:
If a browser does not support an embedded image or sound format, it may use a second software package to open and view the embedded file.

Home Page:
The document that provides a starting point or organizational center for any collection of documents.

Hot List:
List of frequently accessed Web location names and URLs.

Hyperlink:
A word or phrase that is identifiable from surrounding text. Clicking on such text will bring up the destination document that it is "linked" to. Standard hypertext appears a underlined text, in either Blue (untraversed) or Red (traversed).

Hypermedia:
Hypertext that includes or links to other forms of media.

Hypertext:
Text that, when selected, has the ability to present connected documents.

Hypertext Markup Language (HTML):
The standard language used for creating hypermedia documents within the World-Wide Web.

Hypertext Transmission Protocol (HTTP):
The standard language that World-Wide Web clients and servers use to communicate.

Inline image:
A graphic within a hypermedia document that is displayed on the same page as text.

Joint Photographers Expert Group (JPEG):
A compressed format that stores high quality images in relatively small files. Newer than GIF and not as prevalent, they are useful for information rich images.

Motion Pictures Entertainment Group (MPEG):
A motion image format for displaying images and sound.

National Center for Supercomputing Applications (NCSA):
A federally-funded organization whose mission is to develop and research high-technology resources for the scientific community.

Sun Audio (AU):
A popular sound file format.

QuickTime:
A digital video format developed by Apple Computer that integrates synchronized video and audio with compression techniques.

Tags:
Formatting codes used in HTML documents. These tags indicate the structure of a document as it will appear when displayed by a Web browser.

Uniform Resource Locator (URL):
A standardized way of representing different documents, media, and network services on the World Wide Web.

Webmaster:
The administrator responsible for the management and often design of a World-Wide Web site.